This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TVCG.2016.2610422

A Survey on Visual Approaches for Analyzing Scientific Literature and Patents

Paolo Federico, Florian Heimerl, Steffen Koch, *Member, IEEE,* and Silvia Miksch, *Member, IEEE*

Abstract—The increasingly large number of available writings describing technical and scientific progress, calls for advanced analytic tools for their efficient analysis. This is true for many application scenarios in science and industry and for different types of writings, comprising patents and scientific articles. Despite important differences between patents and scientific articles, both have a variety of common characteristics that lead to similar search and analysis tasks. However, the analysis and visualization of these documents is not a trivial task due to the complexity of the documents as well as the large number of possible relations between their multivariate attributes. In this survey, we review interactive analysis and visualization approaches of patents and scientific articles, ranging from exploration tools to sophisticated mining methods. In a bottom-up approach, we categorize them according to two aspects: (a) data type (text, citations, authors, metadata, and combinations thereof), and (b) task (finding and comparing single entities, seeking elementary relations, finding complex patterns, and in particular temporal patterns, and investigating connections between multiple behaviours). Finally, we identify challenges and research directions in this area that ask for future investigations.

Index Terms—Visualization, Scientific Literature, Patents, Documents, Survey

1 Introduction

This article aims at shedding light on interactive visualization approaches supporting search and analysis of scientific articles and patents. We focus on this topic, because there has been an increase in visual approaches for these domains and corresponding tasks during the last few years. This trend can be explained by and correlates with the strongly rising numbers of documents in both domains. From our perspective, the introduction of visualization offers additional levels of aggregation and overview, and helps to deal with bigger sets of these complex documents.

Searches for and analyses of scientific literature and patents typically have high demands on precision, and, particularly, on recall. This distinguishes the tasks that need to be supported in these searches and analyses from those in many other domains. The iterative nature of the analyses, that are typical for complex information-foraging tasks on these documents, render visual approaches particularly suitable for them. Visualization techniques used vary greatly from explanation tools to exploratory methods, but search and analysis goals are comparable in some cases. Straightforward approaches summarize results visually, while advanced ones integrate suitable interaction concepts enabling users to steer algorithms. Immediate updates of visual representations, as a result of such user interactions, close feedback loops and are characteristic of visual analytics approaches.

1.1 Characteristics of scientific literature and patents

We surveyed visual approaches for analysis of scientific articles and patents (with a numerical preponderance of the former, see Section 3). Indeed, scientific articles and patents have a variety of common characteristics which motivated a joint discussion for these two objects of analysis. These common characteristics result in similar search and analysis tasks for both types of documents. Both have a rather similar set of bibliographic information, that we refer to as metadata, attached to them. This includes authors for scientific articles, and applicants for patents. While scientific articles are typically annotated with their date of publication, patents have a much richer set of time-dependent meta data, e.g., filing/application date and many other legal events that may occur within a patent's life cycle.

Another commonality shared by scientific articles and patents are *citations* of relevant previous documents that explain the context a document is embedded in.

Citations and references, however, are not the only source of *relational information*. Patent families, which are based on common (sets of) priorities, are another example. In addition, documents can be linked through metadata, e.g., the same (co-)author. Both document types typically have manually created *classification* schemes associated with them. The ACM taxonomy[1] and the Physics and Astronomy Classification Scheme (PACS)[2], for example, are popular for publications in the respective fields. Depending on the regional office, different classification schemes are in use for patents. These include the Cooperative Patent Classification (CPC)[3] system, as well as former schemes such as the

- *The authors of this paper are listed in alphabetical order*
- *Paolo Federico and Silvia Miksch are with the Institute of Software Technology and Interactive Systems, Vienna University of Technology. E-mail: <lastname>@ifs.tuwien.ac.at*
- *Florian Heimerl and Steffen Koch are with the Institute for Visualization and Interactive Systems (VIS), University of Stuttgart. E-mail: <firstname.lastname>@vis.uni-stuttgart.de*

Manuscript under review.

1. http://www.acm.org/about/class/
2. https://www.aip.org/pacs
3. http://www.cooperativepatentclassification.org/index.html

International Patent Classification (IPC)[4], and the United States Patent Classification (USPC)[5]. Most of these schemes are organized in the form of hierarchical taxonomies.

Geographic information is typically available in the form of regions where a patent is in force, locations of applicants, and locations of institutions authors are affiliated with. These attributes contain structured information on patents and scientific literature and are well-suited to create different visual perspectives on collections of documents. Accordingly, they are often used to create overview visualizations that help stakeholders to get a general understanding of document sets, or, combined with suitable interaction methods, support them in navigating document sets.

The content of both document types has common characteristics as well. While the main information is typically encoded as text, patents and scientific literature are often enriched with additional elements. These include formulas, tables or lists, but also drawings, photos or videos. The use of these elements results in multimodal documents. Both types of documents have technical content, and patents are written with some legal jargon sometimes referred to as 'patentese.' The goal of achieving an unambiguous definition of what a patent covers conflicts, to some extent, with the applicants' demand for broad coverage of legal protection. In combination with the requirements for legal texts and the formal rules for writing patents, this makes patent documents difficult to read and understand for non-experts. Additionally, language processing, mining, and retrieval approaches can lead to results that do not meet the expected quality. Visualization offers means for quickly assessing (intermediary) results, helping users to quickly improve automatic procedures to a satisfactory level of quality.

Scientific documents as well as patents undergo evaluation processes to get published or granted. For both, scientific documents as well as patents, technical novelty is a fundamental requirement for a positive assessment. However, the notion of novelty is different for both types of documents. While it evolves over time for scientific disciplines, and has a generally broader definition, e.g., by also comprising insights from comparative studies, its definition for patents is much more narrow. Rules for 'patentability' are defined in corresponding laws and regulations and differ by country. In general, these rules require that a new 'technical' invention is applicable in an industrial way, and that it must involve an inventive step (c.f. article 33 of PCT). Apart from exhibiting a certain level of novelty, both document types have to fulfill different additional requirements to be 'acceptable'. And although they are formulated differently in practice, tasks such as finding similar works and comparing them are common for both domains. Again, interactive visualization can speed up such tasks.

1.2 Related Surveys

There are a couple of related surveys and other collections of techniques and publications that we want to discuss in this section. Bonino et al. [1] describe the technological state-of-the-art in patent analysis. They introduce patent databases, tasks important for stakeholders, and recent technological innovations. Their survey does not include any visualization approaches or techniques at all, while ours focuses entirely on visualization.

In a more recent patent analysis survey, Abbas et al. [2] identify visualization techniques as an important group of approaches to analyze and understand patent data. While they list some visualization approaches to patent search and analysis, they entirely focus on the patent community. As a consequence, they omit all approaches published to other communities, such as the visualization community. In this survey, all relevant approaches from the visualization community are included and discussed.

Yang et al. [3] review only commercially available text mining and visualization approaches for patents. We have decided to exclude these from this survey, the main reason for this being that available information about them is scarce. Lupu et al. [4] give an in-depth account of current techniques for patent retrieval. While this also includes visualization approaches to retrieval, it does not cover the whole gamut of visualization approaches to patent analysis, as does this survey.

In addition to related surveys on patent retrieval, there are more general publications that focus on certain aspects of scientific literature, often also comprising visualization approaches. Chen [5] gives a broad account of visualization approaches for citations, collaborations, and scientific communities in general. He discusses these approaches in the context of research fronts and their evolution. While we focus on visualization of aspects of scientific publications, Chen focuses on knowledge visualization and the evolution of scientific knowledge in a broader sense.

With her *Atlas of Science*, Börner [6] focuses on mapping approaches, i.e. spatializations of scientific communities often based on citation data. She provides a comprehensive overview including hand-drawn examples. Our focus, in contrast, is broader, comprising all types of visualization methods that provide benefits for the analysis of scientific documents and patents.

Search and retrieval tasks play a vital role for patent and scientific document analysis, and this survey includes multiple visual approaches to support them. Hearst [7] gives a detailed account of these types of interfaces, discussing, among other topics, their design, evaluation, and query refinement techniques. Visual text analysis is another, broader group of techniques relevant for patents and scientific documents. Alencar et al. [8] provide an overview of these techniques. Another comprehensive collection of such approaches that can be searched interactively [9] is available online[6].

1.3 Terminology

With the terms *scientific document*, *scientific publication*, *paper*, and *article*, we refer to articles published in a scientific journal or in the proceedings of a scientific conference. For the patent domain, we use the terms *patent* and *patent application* to refer to documents that are in any stage of the process of patent examination, excluding design patents.

4. http://www.wipo.int/classifications/ipc/en/
5. http://www.uspto.gov/web/patents/classification/
6. http://textvis.lnu.se/

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TVCG.2016.2610422

With the term *visualization* we refer to the disciplines of information visualization, data visualization, and visual analytics. We concentrate on approaches that generate visualizations algorithmically from the data contained in or associated with the documents. Visual representations that were created manually in the form of illustrations or schematic drawings are not considered in this survey whether they are created with computer aid or not. However, there are some approaches where an exact distinction is not possible because they contain a 'constructive' component, either for formulating queries visually or for externalizing derived analysis results. Since these couple data visualization and constructive approaches, they are included in this survey.

1.4 Methodology for selecting publications

As this work aims to summarize the state-of-the-art in visualization of collections of patents and scientific documents, our selection of publications is biased towards papers from the field of visualization or visual analytics. The reason for this is simple: many of the approaches containing a considerable visual contribution were developed by the visualization community. Our goal was to identify and collect all publications from this community that contain a technique, visualization, or approach designed to visualize and analyze either patents or scientific publications. In addition, we have also included all publications that present more general approaches whose application to patents or scientific literature is demonstrated by an example or use case within the publication.

We have collected these publications by listing a core set of relevant publications known to us in advance, or identified through a keyword search. Starting from this core sample, we have followed forward and backward citation links to identify additional relevant publications in an iterative way. In addition to publications from the visualization domain, this process yielded many articles from various other communities. We have included them in this survey if they propose a substantial visualization approach for data from patents or scientific publications that can be used for their analysis. These communities include, but are not limited to information science, bibliometrics and scientometrics, patent research, data mining, information retrieval. We have also included relevant introspective work in publication patterns, behavior, and the history of other scientific disciplines. In addition to following citation links, we have scanned through publication titles and abstracts of conferences and journals that were published during the time of writing this article in order to include the latest developments.

There are certainly patents on how to analyze and retrieve patents, and some of them might consider interactive visualization as well. Apart from these, there are commercial products which integrate a variety of visual techniques and approaches for patents and scientific literature analysis. As we consider the former out of the scope of this survey, and the latter often hard to analyze due to the limited availability of information about commercial approaches, we have decided to not include them in this survey. We thus have decided to concentrate entirely on approaches described in scientific publications.

2 Categorization

We have structured this survey according to two orthogonal aspects: tasks and data types. All of the publications are categorized according to different categories along these two dimensions. This categorization is inspired by Shneiderman's task by data type taxonomy for information visualization [10], but it comprises domain-specific data types and a different task framework. In this section we list and specify the categories and discuss our reasons for each of them.

2.1 Data Categories

We have identified four relevant data types in the domain of patents and scientific literature: **Text**, **Citations**, **Authors**, and **Metadata**. We devoted to each of these data types a category, where techniques for that data type are discussed. The categories are not mutually exclusive: since some techniques deal with more than one data type, we have identified a main category, and one or several secondary ones. Moreover, some visual analytics approaches combine different techniques; in these cases, the techniques are listed in the corresponding categories, and their combinations in the additional **Multiple** category. In the following we describe each data category.

Text: The textual content of a publication is its central component, as it encodes its main scientific contribution and a wealth of additional information. Natural language is used to encode information in multiple parts of a publication, including its title, the abstract, the main part, and its conclusion. Some of the information stored in textual form can be extracted automatically. This includes, e.g., the reasons for each of the citations within a publication, and technical concepts that are being used or advanced. From multiple documents, structural information such as prevalent research topics and their popularity over time can be extracted from the texts. As automatic text analysis is an inherently hard task, it is natural to include human users into the analysis loop to steer the analysis. While some of the visualization approaches are directly designed to support interactive exploration and analysis of a document's contents, others use text data to relate documents and gauge their similarities to create a visualization of the dataset. The approaches in this category vary according to the types of text they use from the publications, including titles, abstracts, and full texts, or any combination of the three.

Citations: Citations are an essential part of research papers and patents. They indicate a certain relatedness between the citing and the cited document, such as a topic similarity, an intellectual influence, a reuse, an extension, or even a confutation. The analysis and visualization of citations can support different kinds of tasks. First of all, citations are understood as a measure of relevance; the mere citation count as well as more complex indices (e.g., PageRank [11], H-index [12]) can be used to assess scientific and technological documents and evaluate scholarly contributions and inventions. Conversely, as a measure of similarity, citations can be exploited to identify related works, cluster them, and map the intellectual structure of research fields. Moreover, as a measure of intellectual influence, citations

	Elem. lookup & comparison	Elem. relation seeking	Synoptic (Patterns)	Synoptic (Temporal patterns)
Text	8	7	20	5 (6)
Citations	2	7	9	10
Authors	2	1	2	7
Metadata	2	5	8	2 (3)

a Approaches for data types: *Text*, *Citations*, *Authors*, and *Metadata* (supporting *Elementary* and *Descriptive Synoptic* tasks).

	Aggregation	Labelling	Composition	Multiple views	Tight integration
Multiple/Connectional	11 (6)	4 (3)	3 (1)	12 (3)	5 (2)

b Approaches for data type: *Multiple* (supporting *Connection Discovery* tasks), by level of integration.

Table 1: Distribution of publications across the task by data type classification: *a* contains the classifications for the single data categories, while *b* contains those for the category of approaches combining multiple data types. The numbers in parentheses are the secondary classifications for a category.

can be used to track the dissemination of new concepts and to identify knowledge flows. Finally, combining the three aforementioned aspects, citations can be exploited to support information organization, management, and retrieval. This category comprises methods that directly visualize links between publications, and methods that exploit citation links to derive information about publications.

Authors: In the context of scholarly literature, an author is generally considered to be a person who has made substantial intellectual contributions to a scientific publication, in terms of both scholarship (conception, design, implementation, analysis, and interpretation of results) and authorship (writing, reviewing, revising the manuscript and approving its submission). In the patent domain, the concept of authorship can be diversified into three different roles: the inventors (who conceive and develop the invention), the applicants (who filed the application for a new patent), and the assignees (who hold the ownership of the patent). The laws are different across countries and patent offices, but inventors are generally natural persons, sometimes coincide with applicants and, as such, can be considered analogous to the authors of scientific publications, in the context of our treatment. This category includes approaches for the analysis of authors, their scientific output, and their collaboration.

Metadata: Metadata comprises data that is typically associated with scientific publications when they are published, such as authors, titles, conference or journal, and year of publication. These attributes are important because they are generally used as an identifier for publications, for example, in reference lists of papers, in libraries, or on-line search interfaces. Further, metadata attributes often associated with publications or patents are categorizations of their topics. These include author-assigned keywords, as often found in conference papers, and classification schemes such as the IPC for patents, or the ACM classification for computer science publications. In addition, we consider all attributes assigned to documents through human labeled metadata. This includes a vast range of data types, such as human assigned topics, or semantic relations between entities in a text. We only include such approaches in this category if the labeling is done without the help of automatic techniques, before or during the visual analysis process. Some of the document attributes mentioned in this category, such as title and authors, have already been described as part of other categories. This is due to the fact that we treat them differently depending on their role in an approach. If an approach exploits a specific method for these attributes (e.g., text mining for titles, social network analysis for authors), then it will appear in the corresponding category (text or, respectively, authors). Conversely, if these attributes are used as document identifiers or multivariate categorical data, the approach will be part of the metadata category. This category also includes approaches that allow document searches by metadata, and ordering them according to their metadata attributes to get an overview of a large collection.

Multiple: Some of the approaches in this survey combine more than one of the data types mentioned so far. They can be categorized into two types. The first type are approaches that loosely couple data types, for example, by providing multiple views on a document data set, each backed by a different type of data. They can, e.g., comprise a scatterplot view based on document text similarities, a graph view of a citation network, and a co-author graph, combining them by techniques such as brushing and linking as part of a user-centric analysis loop. Such loosely coupled techniques will appear multiple times in this survey, in all of the data categories they pertain to. We tried to assign a main category to each of these cases, discussing the approach in-depth in the respective section. Contrary to this loose coupling, approaches that are part of this category closely couple data types. We have created an additional category for them, as they do not treat the different data types in isolation. Presenting such approaches across multiple categories would not accurately account for their often elaborate techniques of combining these data types either visually, or through data mining methods. Examples of approaches in this categories include co-author or citation graphs enriched with content information from the publications, and topical trends that are visually linked to the communities cited by the respective publications. In both cases, combining multiple data types provides a fuller picture of a data set and helps to understand certain findings and phenomena.

2.2 Analysis Tasks

The second dimension for classification we used for this survey is the analysis task. We revised and discussed several possible schemes, and the one that we finally used is derived from the task framework by Andrienko and Andrienko [13]. We used this classification as sub-categories for each of the sections that group approaches by data type. The rest of this section lists and explains the task categories.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TVCG.2016.2610422

Elementary Lookup and Comparison: Approaches in this category are designed to support analyses of single entities. These are, for example, identifying relevant entities from a larger set, or collecting and accumulating information about a certain entity. The entities can be of many types, including, but not limited to, the above listed data types; they include documents, topics, authors, conferences and journals, and many more. Approaches in this category are not limited to visualizations that only depict one single entity at a time, but can also show more than one, enabling comparison between them. The joint characteristic of the approaches, however, is that they support analysis tasks aiming at single data entities. For example, visual document retrieval techniques often comprise a visualization of result sets, while their analysis process is designed to compare results and identify the most relevant documents to a user's search objective.

Elementary Relation Seeking: Another kind of analysis goals are those focusing on relations between entities. The approaches in this category support analysts with identifying entities that satisfy a given relation (e.g., finding publications that cite some of the same literature), or identifying relations between entities (e.g., identifying the content similarity between two publications that have an overlap in their citations). Relations supported by the approaches in this category can be between entities of the same type, or of different types. Examples of the former include the analysis of author collaborations, while examples of the latter include approaches to analyze the relation between documents and the technical concepts they contain.

Synoptic tasks (Patterns): If the analysis is not focused on single entities, but rather on all the elements of a dataset and on the system of relations existing between these elements as a whole, then the approach is part of this category. The approaches therein support the identification and comparison of patterns. An example for such patterns are collaboration circles of scientists: Approaches that allow identification, and exploration (including central and peripheral authors) of such circles are part of this category. Science maps also fall into this category; they show scientific disciplines by grouping publications according to citation links. Depending on the data set used, this results in groups of documents that represent entire disciplines of science and show their patterns of interconnectedness through citations.

Temporal patterns: Among all the approaches supporting synoptic tasks, a number of papers in this survey emphasize approaches that concentrate on the temporal aspect of a data set and facilitate the analysis of its temporal patterns. Therefore, we differentiate approaches that support synoptic tasks involving temporal patterns from the others, and assign them to a specific category. It includes, for example, approaches supporting the analysis of topics interacting with each other and changing over time, of citation trends, of authorship and co-authorship dynamics. The visualization techniques in this category can be split into two subgroups. Some of them map time to space, which can be achieved by introducing a dedicated axis for the temporal aspect of entities. The other type of visualizations map time to time resulting in an animation that conveys the temporal dynamics of a data set. Some approaches within the above categories might also support analysis of temporal aspects of the data. In such a case, if we did not identify insights into temporal phenomena as its main analysis goal, we discuss the approach primarily in another category and only mention it briefly in this one for the sake of completeness.

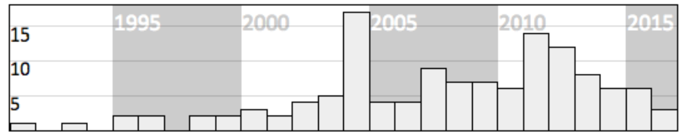


Figure 1: Number of approaches per publication year.

The only data category that does not use the above-mentioned task sub-categories is **Multiple**, which comprises techniques and approaches combining two or more data types, often including temporal aspects. As such, they are aimed at supporting complex synoptic tasks. For this reason, we do not classify them according to the task, but rather according to the way the different data types, analytical methods, and visualization techniques are combined, ranging from a simple composition to a seamless interactive visual analytics integration. With reference to the task framework by Andrienko and Andrienko [13], the elaborate approaches in the **Multiple** category support the most complex synoptic tasks, namely *connectional* tasks (e.g. finding significant connections between phenomena, such as cause-effect relations or structural links). Conversely, the approaches in the **Synoptic tasks** and **Temporal patterns** subcategories of each data category support simpler synoptic tasks, namely *descriptive* tasks.

3 Techniques

Overall, we surveyed 109 approaches for scientific documents and 21 for patents. These approaches were published between 1991 and 2016 (see Figure 1). In this time period, the ratio of techniques addressing scientific documents and patents has been relatively stable. As for data types, the last decade has seen more approaches focusing on text analysis and visualization, while in the previous decade research addressed more authors and citations. Techniques have always combined different data types, but a tighter integration of interactive visualization and automated analysis has emerged in the last decade, in correspondence with the development of the visual analytics paradigm. In this section, structured according to our categorization, we present and discuss all of the surveyed approaches. All techniques and trends can be browsed through an interactive online tool[7].

3.1 Text Data

This section discusses approaches based on analyzing textual content of publications. Some of the following approaches integrate their visualization techniques with advanced natural language processing (NLP). These NLP approaches are used for data preprocessing or directly integrated within the visualization pipeline as an integral part of the sense-making loop. In the latter case, users are typically enabled to steer the automatic analyses of text data in order to adapt and exploit these techniques for their analysis

7. http://www.paperviz.org

needs. We mention these NLP methods if they play an important role in a visualization approach. However, due to a lack of space and because we focus on visualization, we will not explain them in detail. For this, we refer the readers to introductory texts on NLP [14]–[16], and to the original papers listed in this survey.

3.1.1 Elementary Lookup and Comparison

This category mainly contains approaches for document retrieval, an area of research that deals with finding, analyzing, and comparing texts. Users typically provide a query, often a list of search terms, and the system lists the best matching documents, often in order of their relevance. Retrieval techniques used in the subsequent approaches are the vector space model, and Boolean retrieval. While the former statistically analyzes word occurrences to quantify document similarity and relevance to a query, Boolean retrieval just uses information about the presence or absence of search terms within a document. Multiple visual techniques have been proposed to improve the basic retrieval schemes through user interaction, or by giving user additional information besides the bare relevance order.

TileBars [21] is a visualization for Boolean search results that is based on a text tiling technique to split texts into coherent thematic sections. It depicts a visual summary of the distribution of search terms across the sections of a document as a rectangular strip. Through the length of the strip users can compare relative document lengths and differing term distributions in the result set. Nowell et al. [22] present a highly flexible way of organizing search results. Users can organize and explore result sets in the cells of a matrix view whose axes can be set to any aspect of the documents, e.g. author and year of publication. Additional metadata attributes can be included by mapping them to shape, pictograms, label, or color of the respective document icon in its matrix cell. Koch et al. [23] introduce a visual analytics approach for query extension and refinement of Boolean queries. It supports the exploration of result sets through multiple linked views of the distribution of different metadata aspects. The initial query can be extended visually by interacting with these views.

Other approaches allow users to define multiple queries and visualize documents based on their relevance to them. Olsen et al. [24] propose the VIBE system that positions documents on a 2D plane relative to multiple queries. These queries can be defined and positioned freely by users, which results in a highly interactive system for finding and exploring document sets. Scalability is limited, however, as the positions of the documents become ambiguous for four or more queries. GUIDO [25] is a similar method that, other than VIBE, does not support free query placement. It optimizes query position and maps documents according to their absolute distances to queries. Although this theoretically introduces less information loss, the resulting complex geometric forms are harder to interpret. Sparkler [17] uses a different spatialization scheme to facilitate the comparison of multiple queries. It distributes result sets on a circle split into one segment per query. Documents can occur in multiple sets, and their glyphs are colored according to the query, as depicted in Figure 2a. Distance from the center encodes the relevance to a query. To avoid overlaps, documents are spread out radially within the boundaries of the segment. This gives users an overview of the relevance distributions for each query and helps to compare results, e.g., for patent prior art search.

Costagliola [26] presents an approach that also spatializes search results. A user submits a query and the results are laid out according to their textual similarities on a circular area. This area is further extended into a 3D tube by adding a third dimension for the publication time of the articles. The interface supports the standard 3D interactions to counteract occlusion. References between the articles in the set can be displayed as edges on demand.

An approach that is designed for providing an overview of a dataset, rather than retrieving documents, is Document Cards [27]. Each publication is represented by a card that contains descriptive terms as a brief summary of its content and representative images. In addition, the cards provide direct access to individual pages of the documents through interaction.

3.1.2 Elementary Relation Seeking

Chuang et al. [18] present an approach to find topical relations between different university departments based on their doctoral dissertations. The dissertations are reduced in their dimensionality by a method that converts word vectors for documents into lower-dimensional vectors of topics (topic modeling). These topics are represented by weighted lists of words that are extracted from the texts in the dataset. The topic vectors are then mapped to 2D using Principal Component Analysis (PCA). It projects the document vectors onto the two directions in their original space along which they exhibit the highest variance. Such a projection gives an overview of document similarities, but also distorts the original distances. In order to gauge the true distances, users can select a single department and lay out the others in a circular fashion around it, showing undistorted distances. An example of this is depicted in Figure 2b. A method to find and relate thematic clusters in a citation network is presented by Nakazawa et al. [28]. They use topic modeling to group documents thematically and then depict these clusters as nodes of a graph.

Görg et al. [29] help users find relations between biomedical publications based on the biological entities mentioned in their abstracts and index terms, such as genes. The approach helps users find new interesting publications, and enables them to correlate and combine findings about interconnections between entities to generate new insights. A more general version of this approach [30] allows analyzing publications from any domain. It combines various text analysis methods including clustering of documents by topic and named-entity extraction and correlation with metadata attributes of the documents. Riehmann et al.'s [31] approach is also designed to relate and compare text from different documents to explore cases of suspected plagiarism. The visualization is based on a bipartite graph, linking text passages in the suspicious documents with possible sources.

Rexplore [32] is a web-based system for search and faceted browsing of publications that features a node-link graph connecting similar authors. Similarity is quantified by comparing the trajectories of authors through research

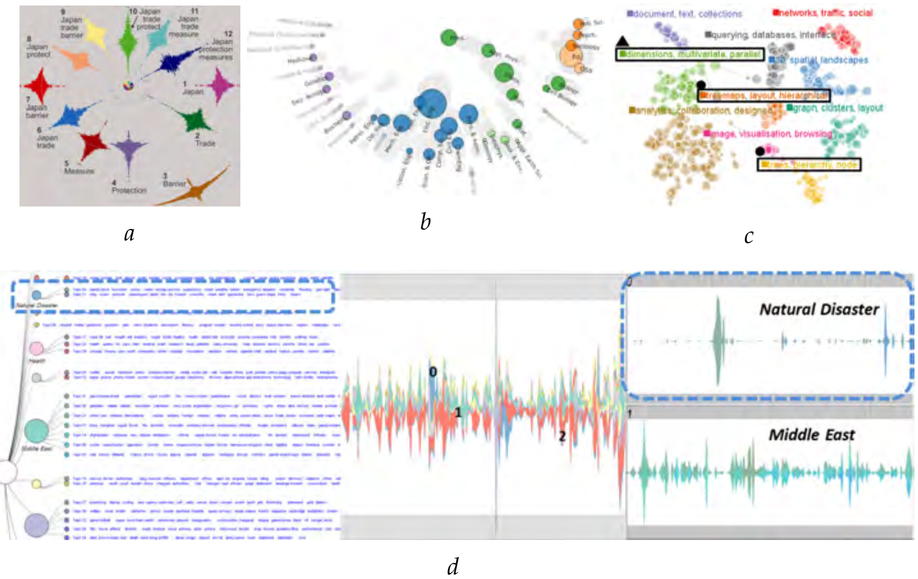


Figure 2: Examples for visualization of text data, by task: *a* elementary lookup and comparison (in [17]); *b* elementary relation-seeking (in [18]); *c* patterns (Utopian, [19]); *d* temporal patterns (HierarchicalTopics, [20]).

topics extracted automatically from the document contents during a pre-processing step. PaperLens [33] groups papers by research topics. The topics are automatically extracted from the text and then depicted as barcharts per year. In addition, it depicts rankings of the ten most often cited authors per year, and allows users to find connections between authors in a co-author graph.

3.1.3 Synoptic Tasks (Patterns)

In addition to supporting relation analysis, Görg et al.'s [30] approach presented in the previous section also contains a visualization that depicts document similarity patterns on a 2D plane. This provides users with information about similar group patterns of documents in a data set. The Vx-Insight [34] system is specifically modeled to visualize such automatic groupings of documents. It bases them on one of several notions of similarity selected by the user, including textual similarity and citation links between papers. The resulting 3D visualizations follow a map metaphor representing dense areas as hills and areas of lower density as valleys. Users can zoom and rotate the resulting terrain, and the valleys are labeled with representative terms to assist navigation. Analysts can thus overview a data set and see its main topics and their distribution patterns. In-Spire [35] also enables document set spatialization. It provides two different visualizations, the *GalaxyView* and the *ThemeView*. The former is a scatterplot of documents in 2D, while the latter uses a 3D metaphor with mountains for dense and valleys for less dense areas.

The GistIcons [36] approach produces a circular histogram of terms for each document. To make the resulting, individual shapes comparable, the terms are grouped by concepts. Based on the visual similarity of the shapes, users can identify topics and groups of similar documents. Wu et al. [37] achieve a similar effect using traditional word clouds, presenting a new algorithm to optimize them. The Termite system [38] is designed to give insight into automatic topic modeling results. It consists of a term-topic matrix that includes the subset of most distinguishing terms between the topics. The relevance of a term for each topic is depicted by circles of varying size. This provides users with an overview of the topics in a document set and their meaning. Chuang et al. [39] use this technique for the exploration of topics in a set of PhD dissertations. They combine it with a citation graph along a time line that visualizes the influence from a cited to a citing paper as topic flows. The documents, depicted as nodes of the graph, are sized according to their overall influence on others. This shows historic developments over time and helps to identify highly influential papers.

ParallelTopics [40] is another approach designed to analyze and explore topic modeling results. It includes multiple views, including word clouds for each topic, and a streamgraph to show the temporal dynamics. In addition, a scatterplot gives information about the number of topics each document contains, and a parallel coordinates view gives a detailed account of the topic distribution for selected documents. Jiang and Zhang [41] also use a topic scatter plot to depict topic similarity. They combine it with a Sankey diagram that shows topic evolution in a dataset over time. Gretarsson et al. [42] present a web-based topic exploration approach that shows topic modeling results as node link diagrams and lets users explore their connection to publications and university departments. The iVisClustering technique [43] uses topic modeling that can be steered by users to visually classify documents based on its results. Österling et al. [44] combine topic modeling with density estimation to extract cluster structure from a document dataset. Their visualization is a 3D spatialization technique that represents document clusters as hills or islands on a plane, depending on user preference. The authors show how their technique can be applied to a patent set spanning eight different IPC categories. Oelke et al. [45] focus on document sets of up to three classes. These classes are depicted by splitting up a rectangular area into up to three subareas. Circular document coins that contain word clouds of topics extracted from the underlying documents are placed within or at the border of areas, specifying the affinity of a topic to one or multiple classes. The authors demonstrate their technique by comparing the topics of papers from InfoVis, SciVis, and Siggraph.

Choo et al. [19] propose UTOPIAN, a system that bases a 2D mapping of documents on topic extraction results. Through this 2D scatter plot, on which clusters are labeled with keywords, users can interact with the topic modeling algorithm. Possible interactions comprise topic refinement by modifying the weight of keywords in a topic, merging of topics that are similar, splitting of topics, and creation of topics based on user selected documents or keywords. The authors show how their approach can be applied to a set of publications from InfoVis and VAST. An example mapping with this dataset is depicted in Figure 2c. A different approach to grouping documents into clusters is presented by Kohonen et al. [46]. They create a 2D map of a large number (about 6 million) of patent documents. The approach features a search interface to highlight documents on the map by specifying keywords.

Maps are a popular visualization metaphor for scientific disciplines or publications. Many of the mapping approaches are based on either citation or co-author data, but there are examples that create their map based on textual data. Fried and Kobourov [47] create a map of computer science publications based on titles from the DBLP database[8]. Their algorithm extracts keywords, and links them based on word co-occurrences. Users can activate a heatmap that highlights certain areas of the map. Thus, profiles of researchers, research institutions, or conferences can be depicted relative to the entire map. Skupin [48] creates a map based on contents of conference abstracts. The resulting groups are clustered a second time into a hierarchy from which maps of various granularity can be created. Skupin [49] presents his entire pipeline into which he plugs multiple other clustering methods, discussing and evaluating the resulting maps.

3.1.4 Temporal Patterns

Depicting temporal developments and dependencies is part of some of the above approaches in the entities [26], relations [32], and patterns [33], [34], [39], [40] category. Except Davidson et al. [34], who produce visualizations of multiple time slices, all other approaches map time to space by introducing an additional time axis in their respective visualizations.

In Mane and Börner's [50] approach, temporal dynamics play a central role. They extract important terms from titles and keywords of biological articles and plot their occurrences over time. This allows the detection of bursts caused by sudden interest in a particular theme. Term co-occurrences can also be depicted as a graph, encoding the temporal dimension by the color of the nodes.

Creating and visualizing a hierarchy of topics is explored by Dou et al. [20]. They create the topic hierarchy through a combination of topic modeling and a hierarchical clustering algorithm. The resulting tree of topics is depicted in a node-link fashion and can be modified and adapted according to the user's analysis goal. Topics associated with each node can be explored in a streamgraph view that depicts their development over time, as can be seen in Figure 2d. Streamgraphs [51] are a popular visualization technique for time-dependent data that maps time to a spatial dimension. Ahmed et al. [52] use a 3D approach similar to streamgraphs to visualize clustering results on a data set of publications (InfoVIS 2004 dataset). Thematic clusters are depicted over time as worm-like 3D structures of varying thickness. Citations between papers in different clusters are depicted by edges between time points of the streams. To analyze topics in patents, Ankam et al. [53] also use topic modeling in combination with a streamgraph visualization. Their approach includes a radar chart that provides information about the distribution of topics across IPC classes.

3.2 Citations

In order to support diverse analytical tasks, documents and citations between them can be modeled as different kinds of bibliographic networks [59]. In the simplest case of a direct citation network, vertices represent documents and edges represent citations between them. Several types of indirect relations can be derived from the direct citation. Bibliographic coupling, for example, is defined as the relation between two documents that cite one or more identical documents. Co-citation is defined as the relation between two papers that are cited by at least one common document. Longitudinal coupling occurs between two papers if there is a path of citations connecting them in two or more steps.

3.2.1 Elementary Lookup and Comparison

A well-known visualization, based on an organic metaphor, is proposed by Mackinlay et al. [60]. Their search and retrieval system depicts result sets for user queries as a pyramid of bibliographic entries. A single document can be selected, and its citations are depicted as a "butterfly" that consists of two "wings" that list their incoming and outgoing citations. Users can navigate these citations to find more potentially relevant publications.

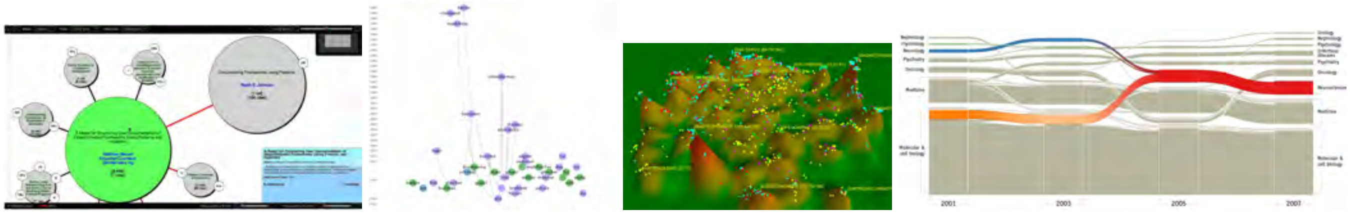
CircleView [54] is a visualization technique showing a focus paper and two levels of its citation network (Figure 3a). The visual representation is based on a node-link diagram with a radial layout, the focus paper being in the center. Visual variables (e.g., color, node size, edge thickness) are used to encode bibliometric measures such as the number of outgoing references and incoming citations, while other metadata (publisher, pages) are shown on-demand by a tool tip. By clicking on an adjacent node, the user can switch the focus paper and navigate the citation network step by step.

3.2.2 Elementary Relation Seeking

A first attempt to provide an overview of a direct citation network was Garfield's historiograph [61], a node-link visualization whose layout is arranged along the time axis. It exploits the fact that citations links are directed backwards in time (i.e. recent papers only cite older papers). Besides this feature, the visualization does not use any other layout

8. http://dblp.uni-trier.de
(All URLs in this survey were accessible on September 10, 2016)

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TVCG.2016.2610422



a *b* *c* *d*

Figure 3: Examples for visualization of citation data, by task: *a* elementary lookup and comparison (CircleView [54], [55]); *b* elementary relation seeking (CiteNetExplorer [56]); *c* patterns (VxInsight [57]); *d* temporal patterns (in [58]).

optimization and does not scale well with the number of papers. The overview is thus limited to the most relevant papers only, by filtering out the papers whose citation count is below a certain threshold. CitNetExplorer [56] features a re-implementation of historiographs, tailored for visualization of longitudinal coupling links (Figure 3b).

While a node-link diagram is the most common visualization for direct citations, other network visualization techniques have been applied to this kind of data. Henry et al. [62], for example, aggregate papers by conference, and show inter- and intra-conference citations by a matrix-based visualization. Aris et al. [63] group papers by research fronts, i.e. areas of significant activity in a certain time period. Research fronts are then visualized by multiple scatter plots (year by citation count), with inter- and intra-front citations draw as overlaid lines. CiteGraph [64] also uses scatter plots to visualize citations in a bibliographic collection. One axis is always devoted to citation statistics, while the other axis and additional visual variables encode other bibliographic data.

Small [65] focuses on co-citation as a similarity measure to draw an early map of science. The huge data set is reduced by sampling, and then it is hierarchically clustered by the co-citation count up to five levels. Finally, the layout for each cluster in the hierarchy is computed by triangulation, a fast algorithm for multi-dimensional scaling supporting local optimization. Different other approaches have been proposed to enhance a similar science map; for example, filtering the data set by a citation count threshold and applying a force-directed layout [66].

3.2.3 Synoptic Tasks (Patterns)

The adoption of adequate methods for decreasing the number of data items and computing a good layout is a common problem of science maps. These maps aim to provide a meaningful overview, keep the overall structure, and support the identification of patterns.

As for direct citations, Delest et al. [67] observe that, since citation links are only directed backwards in time, there are no cycles, i.e. the citation topology is a direct acyclic graph (DAG). They introduce a metric that takes into account the global branching structure of the DAG, and exploit it to optimize the coloring, the label drawing, and the rendering order. Moreover they use this metric to cluster the DAG, by aggregating nodes into super-nodes. A more sophisticated approach to highlight the structures of the science map and visualize information flows between disciplines is proposed by Rosvall et al. [68]. They consider a network of 6 million direct citations aggregated at the level of 6 thousand journals, and then apply a random walk algorithm which identifies the fields and also highlights information flows between them. The resulting node-link visualization provides details on specific disciplines and sub-fields, as well as general trends (e.g., information flowing from basic sciences to applied sciences). The VEGAS system [69] demonstrates an influence graph summarization for citation networks: a matrix-decomposition algorithm clusters nodes according to topological similarity and reachability, and thus it highlights flow-based citation patterns. The VxInsight system [57] for patent visualization exploits a linear combination of citation and co-citation counts as a similarity measure, and applies a force-directed algorithm to draw a 3D landscape of a patent data set (Figure 3c). Users can then define searches on the whole data set and the results are depicted as an overlay of points on the map.

As for co-citation networks, besides the basic reduction method consisting in filtering nodes by the overall number of citations, one can take into account the number of times two papers have been cited together as their link weight and then apply more advanced algorithms for weighted networks, such as minimum spanning tree (MST) and pathfinder networks (PFNET) [70]. An MST of a network is a subgraph that is a tree, connects all vertices, and has minimum weight. A PFNET can be understood as a generalization of MST; it is a link reduction method which essentially filters out links that are not on shortest paths. A comparative application of MSTs and PFNETs to co-citation networks shows that MSTs are computationally more efficient and tend to form clusters that can be intuitively perceived as hubs and authorities, but PFNETs are actually able to better preserve structural patterns [70]. Several similarity or proximity metrics have been proposed as alternatives to the co-citation count for feeding the link reduction algorithms. Noel et al. [71], for example, consider the citation correlation (more properly, the Pearson product-moment correlation coefficient) and compare it with the co-citation count as similarity metrics for computing an MST-based visualization of influence networks. Their results show that, despite the greater computational complexity of the citation correlation, the citation count better supports the identification of patterns of influential documents. Zhang et al. [72] describe another approach for finding patterns in a co-citation network exploiting a Frequency Pattern (FP) tree, a compact prefix data structure for storing events that occur frequently together. They combine an FP tree and a word tree visualization to represent a paper-reference network

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TVCG.2016.2610422

and its co-citation counts. A comprehensive approach to visualize bibliographic networks is proposed by Brandes et al. [73]. Their similarity measure is based on a combination of co-citation and bibliographic coupling, and it is used for computing a modified Laplacian layout. The results are visualized as a landscape visualization, where the surface elevation represents the scholarly relevance, and papers are visualized as houses on the landscape. Van Eck and Waltman [74] present a system to view different types of co-citation networks.

3.2.4 Temporal Patterns

Citations are an inherently temporal phenomenon (newer papers cite older papers), thus many techniques discussed in the previous sections deal to a certain extent with temporal aspects. In the section we present techniques which focus on time as the prevalent part of the analysis and visualization process. As for the analysis, Chen [75] observes that when sampling a large data set that spans over decades, a static threshold for citation or co-citation counts can lead to unbalanced results, because the average number of citations and references can change over time. Hence, he proposes a progressive approach: before applying link reduction algorithms such as MST or PFNET, the data set is partitioned into time slices, then any time slice is filtered by applying an adaptive threshold and pruned by MST or PFNET. Finally the slices are merged back into a single data set for the layout computation. The resulting map is static, in the sense that the data has been flattened over time, but it has also been temporally normalized.

Other approaches take temporal aspects explicitly into account to design the visual encoding, for example, by including a time line. Herr et al. [76] present a visualization of scientific articles in the domain of physics. They do not reduce the large data set by filtering, but rather aggregate the documents by journals and journal volumes. The resulting clusters are visualized in a box layout, with the PACS (Physics and Astronomy Classification Scheme) classification along the vertical axis and time along the horizontal axis. Citation flows are overlaid on top of the boxes, by using an edge bundling algorithm, to show citations patterns over time. Citeology [77] builds upon the already mentioned historiograms [61] and computes the layout of a citation network along a time axis. It does not reduce the data set by filtering or aggregation, but rather exploits several interaction techniques to cope with the large size and to enable effective exploration.

Another possibility to visualize time is animation. GraphaEL [78] is an animated node-link visualization for dynamic citation networks; it features an evolving graph layout that enables the identification of changes while preserving the user's mental map. The rising landscape technique [79] animates two visual variables to show the evolution of a co-citation pathfinder network rendered as a 3D landscape. The transparency of nodes and links encodes the publication status (pre-print, published, and cited/co-cited, respectively), while the third dimension is used to visualize the citation count of each paper, as a colored bar rising from each node.

Other techniques deal with changes over time by using small multiples. The seminal work by Small [80], for example, visualizes the evolution of co-citation networks by slicing the data set and juxtaposing several contour maps (obtained by a combination of a containment technique and a multi-dimensional scaling algorithm). Shibata et al. [81] use a small-multiples approach to visualize the evolution of a direct citation network sliced, clustered, and then drawn by a force-directed algorithm for either a stable or a dynamic layout. Rosvall and Bergstrom [58] propose a sophisticated approach to identify structural changes in a citation network by considering temporal aspects not only in the sampling or the layout phase, but also in the clustering phase. They address the problem of tracking clusters over time which includes identifying split and merged clusters. To solve this problem, they apply bootstrap sampling and simulated annealing algorithms and compute the significance of clusters across splitting or merging. Then, the evolution of citation flows is visualized by an alluvial diagram (Figure 3d).

Temporal aspects can be also addressed by proper interactions. Abello et al. [82] observe that providing both overview and detail on a dynamic citation network is a challenge, and small changes can be drowned out by larger ones. Thus, they provide a degree-of-interest specification by which the user can identify salient changes at the desired scale and importance. The approach by Chen et al. [83] aims at delineating the citation impact of scientific publications. It measures the impact by an adapted version of the H index [12], while a decision tree algorithm identifies emerging topics. Users can review extracted topics and explore keyword bursts through a CiteSpace [84] visualization.

3.3 Authors

In many analysis and visualization approaches, authors are not considered mere metadata, but rather separate entities linked to the scientific documents or, respectively, the patents.

3.3.1 Elementary Lookup and Comparison

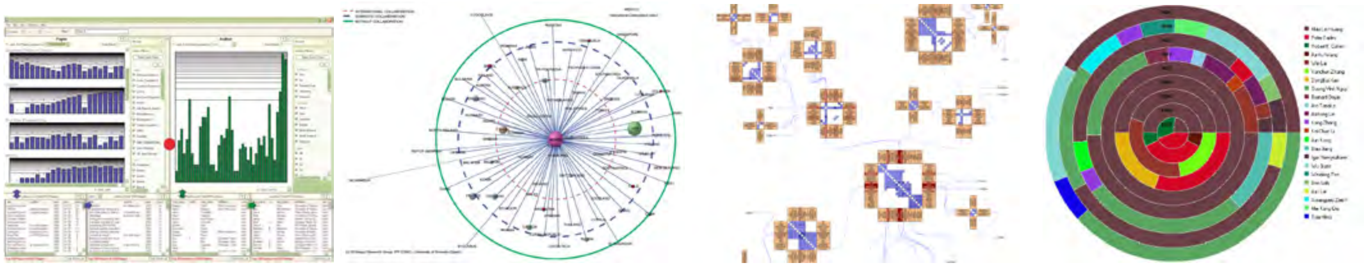
A common approach is to model authors and documents as nodes of a bipartite graph, connected by authorship links. Thus, the visualization can exploit specific graph drawing techniques, such as anchored maps [89]. An anchored map is a node-link diagram for bipartite graphs where the nodes of one set are fixed as anchors with a radial layout and the nodes of the other set are arranged with a force-directed layout. In this approach authors and papers can be alternatively displayed as anchors.

NetLens [85] is an alternative approach that supports the iterative exploration of content-actor network data. It also uses a bipartite graph as a data model, but it is not based on a node-link visualization. Conversely, it features coordinated multiple views based on bar charts, showing the distribution of papers and authors over all the available attributes, for example topic, country, or year (Figure 4a). The system supports iterative filtering, with a dual-mode data flow that allows switching the focus from papers to authors and back.

3.3.2 Elementary Relation Seeking

A relevant relation between authors is co-authorship, modeling the collaboration among researchers or inventors.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TVCG.2016.2610422



a *b* *c* *d*

Figure 4: Examples for visualization of author data, by task: *a* elementary lookup and comparison (NetLens [85]); *b* elementary relation seeking (in [86]); *c* patterns (NodeTrix [87]); *d* temporal patterns (in [88]).

Chinchilla-Rodríguez et al. [86] propose an approach to visualize the scientific collaboration at a given level of aggregation, such as institutional, regional, or national. Their visualization features an egocentric node-link diagram with a radial layout (Figure 4b): the central node represents a given country, other nodes represent countries whose researchers collaborate with the researchers of the given country, and links represent collaboration ties. The diagram is enriched with two measures: collaboration rate (node size) and impact factor (distance from center).

3.3.3 Synoptic Tasks (Patterns)

NodeTrix [87] can visualize a co-authorship network with a hybrid approach: intra-community relationships are visualized as adjacency matrices while inter-community relationships are visualized by the means of node-link diagrams (Figure 4c). It enables the visual identification of three co-authorship patterns: the cross pattern (a research group in which a central researcher collaborates with all the others), the block pattern (all researchers collaborate with each other), and the intermediate pattern (many researchers collaborate but there is a prominently central researcher).

Ichise et al. [90] propose an approach to filter out unimportant links and visualize only tight communities in a co-authorship network. The approach is based on three different community mining algorithms whose parameters can be interactively adjusted by users.

3.3.4 Temporal Patterns

While the approaches described in the previous sections are limited to the visual analysis of authors in a static fashion, there are also techniques that focus on the dynamics, i.e., the change of productivity and collaboration patterns over time. Kutz et al. [91] focus on the trends of both the productivity of single inventors and the diversity of their patent portfolios. The approach features tree maps arranged along a time line. Each tree map represents the class distribution of patents granted to an inventor in one year. Keim et al. [92] visualize the dynamics of collaboration by means of two techniques: PaperFinder and InterRing. PaperFinder features a 2D layout, with the authors along the y axis and the time line along the x axis. The papers, depending on their keywords, are assigned to categories encoded by colors. Thus, the visualization shows the development of topic and authors over time. InterRing is a radial, space-filling visualization technique. It encodes years into circular sectors and arranges co-authors along the radial axis, still using colors for categories. Huang et al. [88] also propose the use of InterRings to visualize the collaborations of individual researchers over time, but they map time to the radial axis and introduce an algorithm to compute the weight of each co-author's contribution, visually mapped to the angle (Figure 4d). Shi et al. [93] apply a 1.5D visualization technique to egocentric co-authorship networks. It features a combination of two layout algorithms (force-directed and radial), complemented by trend glyphs.

Animation can be also useful to visualize the dynamics of a co-authorship network. GraphDiaries [94], for example, provides a node-link visualization technique that exploits staged animated transitions and highlighting to support the identification and tracking of changes.

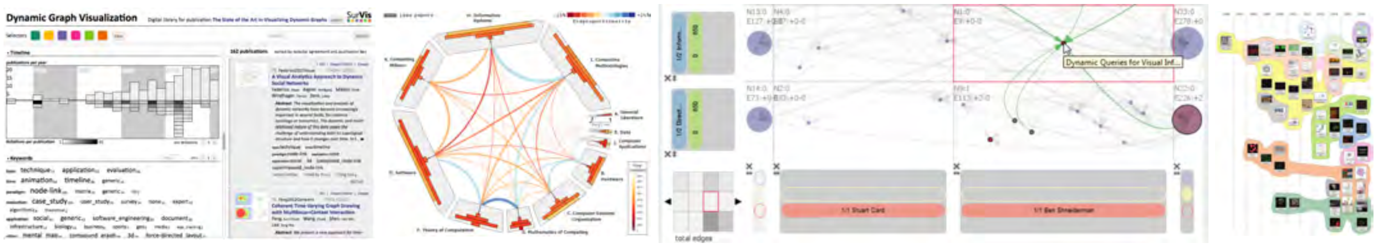
The evolution of co-authorship relations over time can be modeled and visualized as a temporal multi-dimensional network. This is demonstrated by the comprehensive application examples of the Orion system [95]. It features several visualizations, such as node-link diagrams and adjacency matrices. Moreover, it computes authors' centrality measures, both static (visually encoded into scatter-plots) and dynamic ones (encoded into line charts). Kurosawa and Takama [96] propose a visualization system to analyze collaboration between researchers. The system features the visualization of a co-authorship network as a node-link diagram. Each node represents a researcher and is drawn as a radial glyph, encoding the research topics and the temporal trend of the researcher's publications.

3.4 Metadata

This section discusses approaches that are primarily based on metadata of documents.

3.4.1 Elementary Lookup and Comparison

Giereth et al. [101] visualize annotations in patents by distorting the text and enlarging annotated lines. The approach supports semantic annotations that are visualized by node-link diagrams as text overlay. SurVis [97] is a web-based method designed to support survey authors. Figure 5a shows a screenshot of the system. It is based on document metadata including keywords assigned by users, and helps to structure and explore a set of publications. This is achieved by allowing users to create multiple selectors that help to structure the set. In addition, clustering of publications based on metadata is included in the approach.



a *b* *c* *d*

Figure 5: Examples for visualization of meta data, by task: *a* elementary lookup and comparison (SurVis [97]); *b* elementary relation seeking (RadialSets [98]); *c* patterns (in [99]); *d* temporal patterns (BubbleSets [100]).

3.4.2 Elementary Relation Seeking

We introduced Görg et al.'s [30] approach in the text category, but it also supports relating documents through metadata such as year, author, and keywords. We thus mention it here again. Hascoët and Dragicevic [102] present a technique for the interactive visualization of multi-layered graphs. They apply it to a set of InfoVis publications from 1989-1993, defining each year as a separate layer indicated by a distinctive color. In addition, keyword bursts are extracted and linked to the respective papers. This helps to explore the occurrence of popular topics and their dynamics over the years.

RadialSets [98] is an approach for visualizing the overlap between sets of entities. It is based on a radial layout of different sets grouped by data attributes. The overlaps between pairs or amongst larger tuples of sets are depicted as arcs between them. An example of this is shown in Figure 5b for pairwise overlaps. The authors demonstrate their technique with a set of $50,000$ ACM publications grouped by their ACM classifications. For each set, the distribution of publication time is depicted as bar charts. In Figure 5b, the size of the overlaps is encoded by the thickness of the arcs, while the color encodes *disproportionality*, a measure for the statistical significance of the size of an overlap.

Guo et al. [103] propose a visualization for the connectivity of rat brain areas based on data extracted manually from relevant publications. It depicts a radial ordering of coarse brain areas as arcs of a circle with edges between them showing connections.

Nesbitt [104] visualizes relations using the metro map metaphor. Visualizations are constructed by interpreting entities as stations, and relations between them as rail tracks. Relation types are encoded through different track colors. The author presents an example that shows the interconnection of ideas and themes within a PhD thesis.

3.4.3 Synoptic Tasks (Patterns)

The INVISQUE system [105] supports interactive visual analysis of document search results. It is designed for users that have only a vague idea of what they are looking for and supports them in exploring sets of potentially relevant entities to refine their search objectives. Documents are depicted as cards that contain metadata and other information about it. The approach facilitates interactive exploration of a 2D document projection. Documents similar in certain aspects, e.g., in content or metadata can be highlighted.

PivotSlice [99] is based on a 2D matrix visualization. It depicts the distribution patterns of data attributes based on Boolean queries constructed by users. Users can explore each matrix cell and view links between documents (e.g., citation links). The authors demonstrate their system using the InfoVis 2004 contest dataset, as depicted in Figure 5c.

An approach to visualize set-valued attributes of patent documents is presented by Wittenburg et al. [106]. They are depicted as stacked bar charts that show the relative frequency of a specific value, and co-occurrences with other attributes. Wittenburg and Pekhteryev [107] present a comparable technique designed for hierarchical attributes.

Another matrix-based approach for document search results is presented by Shneiderman et al. [108]. The axes of the matrix can be freely configured, and the approach allows highlighting cells based on specific attributes. In addition, hierarchical attributes are supported, whose level can be changed interactively.

Sallaberry et al. [109] present a system to explore contents of publications. Users can analyze documents relevant to specific, user provided, sequential biological patterns. The publications are automatically retrieved and then organized radially around the center of the display. Distance from the center increases with decreasing relevance of the publication.

Other approaches use graph or graph-based visualizations to show patterns. For the IPC class usage frequencies, Giereth et al. [110] use a treemap. On top of each cell that represents an IPC class, a 3D bar chart shows its usage frequency. In addition, co-classification relations can be displayed using 3D edges and edge bundling on the depicted IPC landscape.

Börner et al. [111] use a similar technique in 2D that links citations in patents directly to the taxonomy hierarchy. The approach also includes a notion of document similarity based on the taxonomy.

Perer et al.'s approach [112] is designed to mine interperson relationships in organizations. Analyses are based on data available within the organization, such as co-authored documents or other collaborative projects. Although it is not designed for scientific literature, we decided to include it here, because it is closely related and could be easily adapted to model and analyze scientific collaboration. The approach automatically creates models of personal relationships visualized as an interactive graph that users can browse and filter.

3.4.4 Temporal Patterns

The matrix approaches for search results whose axis can be freely configured [99], [108], also allow temporal exploration of document sets. Users can configure the matrix to map a temporal aspect to an axis, e.g., year of publication for the documents. In addition, the technique by Hascoët and Dragicevic [102] presented above encodes publications years as layers of a graph allowing for temporal exploration of documents and their correspondence to keyword bursts. Bubble Sets [100] is a technique to show mutual relationships among a set of data objects. It is a set visualization that groups instances into one or multiple sets by including them in a colored bubble. The objects can be grouped or organized according to different aspects, e.g., along a timeline. Figure 5d shows the development of different topics over time, providing direct references and icons for each of the publications. Additional information, such as the publication abstracts, is accessible on demand. An approach to track the community development in a patent domain over time is presented by Chen et al. [113]. The communities are extracted by creating time slices of a patent dataset, and clustering citation graphs within each of the time steps. These clusters are linked across multiple time slices based on the overlap between them. The authors combine this technique with a static visualization approach for the communities, depicting them as circles on a scatterplot. Patent numbers are indicated by the size of the circles which are positioned according to their last year and their first year of appearance. The temporal dimension is mapped to the horizontal axis of the scatterplot.

3.5 Multiple

In this section, we discuss systems, techniques, and approaches that combine several data types from the previous sections. These comprise text, metadata, authors, and citations, spanning from simple cases of juxtaposition or sequential application to tighter levels of integration, supporting connectional synoptic tasks.

3.5.1 Aggregation by metadata

A common way to combine two data types is the use of one data type to perform an aggregation into groups, and then the application of a specific visual analytics technique for the other one. In the previous section we have presented examples of authors aggregated by their affiliation or geographic location [86], and citations aggregated by journals or conferences [62], [68], [76]. Common set visualization techniques have been applied to metadata, such as the already mentioned RadialSets [98], or LineSets [118]. The latter technique visualizes a co-authorship network as a node-link diagram, and each group of authors working on the same topic as an overlaid continuous line. A well-known specific example of aggregation by metadata is the author co-citation analysis (ACA) [119]: co-citations between documents are aggregated by authors, and the resulting network is then processed by usual techniques for citation analysis and network visualization, with authors as nodes instead of documents. White and McCain [119] compute several metrics and show them as a bar plot; they also draw a 2D spatialization of the so-called authors' landscape. Lin et al. [120] show how SOMs and PFNETs can be applied to these author co-citation networks. Chen and Paul [114] obtain a 3D visualization by applying PFNETs to both author co-citations and author citations (Figure 6a). Li et al. [121] present a node-link diagram of citations between patents with three different types of aggregation (by institution, country, or technology field), to analyze patterns of knowledge transfer. Windhager et al. [122] model a patent data set as a temporal multivariate network and propose aggregation and projection methods that enable the visual analysis of rising inventors and emerging technologies. Nagel et al. [123] demonstrate an interactive table top visualization of a co-authorship network as a node-link diagram, aggregated by authors' affiliations and spatialized by their geographical positions. Rosvall and Bergstrom [124] present a method based on random walks to automatically organize a graph into multiple levels; they apply it to citations aggregated by journal, obtaining a map of science with a multi-level hierarchy. Honkela et al. [125] draw a SOM of documents and not only show papers aggregated by authors and affiliations, but also complement these bibliographic data by visualizing additional data such as funded research projects. Jusufi et al. [126] use k-means clustering to aggregate document into topic nodes and then depict co-authorship relations as edges between them. Heimerl et al. [117] visualize citation trends along a time line; the standard citation count is complemented with the citation entropy, a measure of diversity computed from publication venues extracted from metadata.

3.5.2 Labels extracted from text and metadata

Another common way to combine data types is the use of metadata for enriching existing visualizations, in particular for labeling clusters obtained by applying specific techniques to the main data type (e.g., for labeling paper co-citation clusters or keyword co-occurrence clusters). Representative keywords can be extracted from metadata manually (like in [127] and [128]), or automatically. Usually keywords are extracted by computing the metadata profile of a cluster and taking the most frequent words. In [65], for example, labels are based on a frequency analysis of article titles and journal category names. Similarly, Van Ham [129] introduces a clustering and layout algorithm for citation and co-authorship graphs, and uses keywords extracted from metadata profiles to label the clusters. Sharara et al. [115] present a system to compare uncertain graphs, featuring multiple views (tabular, matrix-based, and node-link, Figure 6b). In particular, they show an example where the uncertain items to be compared are the labels obtained by two different algorithms. One is based on the text analysis of single documents, and the other one is based on the text analysis of neighboring documents in the citation graph. In CiteScape II [84], clusters of co-cited papers are also labeled by extracting most frequent terms from word profiles of the titles of citing papers. Moreover, temporal bursts of citations are taken into consideration to properly label emerging topics.

A Document Card [27] can be also understood as a combination of labels summarizing the main text (see also Section 3.1), and image miniatures extracted from the document and re-packed into a single-page layout.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TVCG.2016.2610422

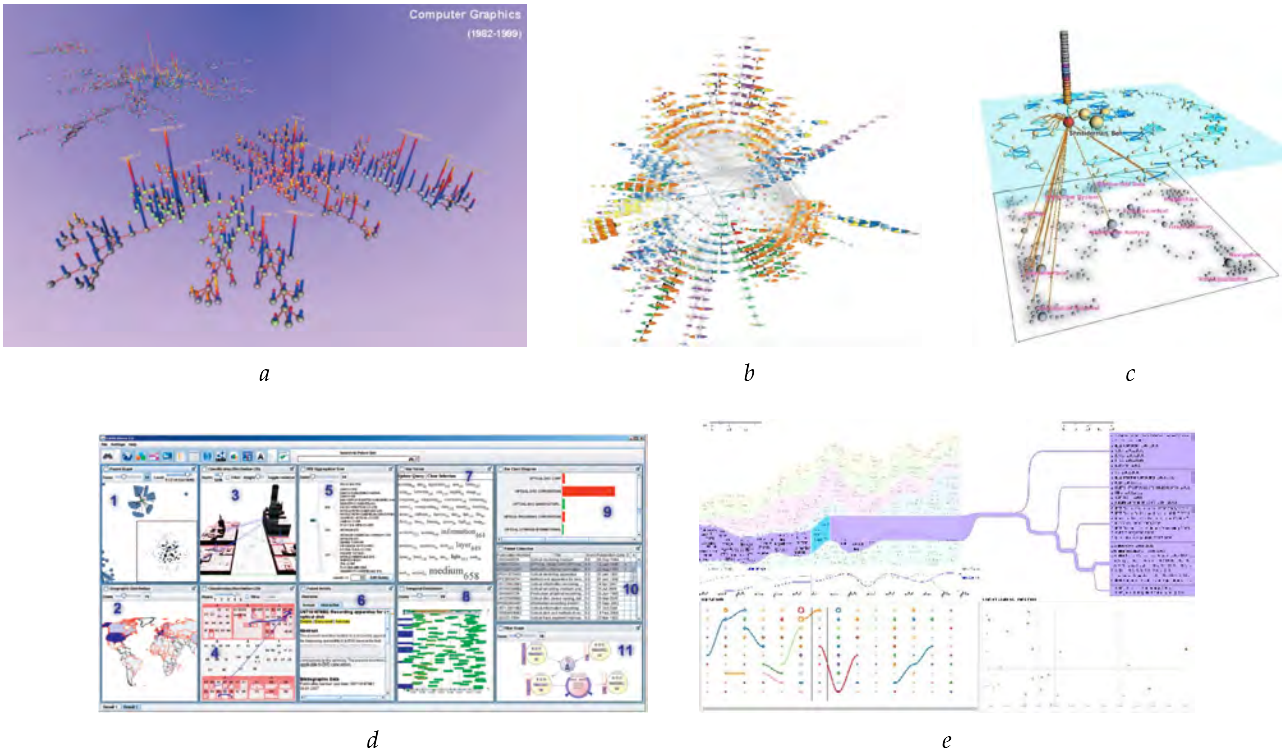


Figure 6: Examples for visualization of multiple data: *a* aggregation (in [114]); *b* labeling (in [115]); *c* visual composition (BiblioViz [116]); *d* multiple views (PatViz [23]); *e* tight integration (CiteRivers [117]).

3.5.3 Visual composition

Visual composition enables the display of different data types within a single visualization. The analysis system for scientific literature by Shen et al. [116] features a 2.5D visualization (Figure 6c): relations between entities are depicted as node-link diagrams with different layout algorithms (radial, force-directed, SOM), and several layers can be stacked on top of each other in parallel to present a coherent view. For example, two node-link diagrams can be stacked on two parallel planes, the one showing authors and co-authorship, the other one showing papers and citations. Additional inter-plane links connect authors with their papers.

Tyman et al. [130] propose a 3D scatter plot where papers are positioned along three axes (keywords, citations, and years) and colored according to the topic classification. Moreover, along one facet of the 3D scatter plot, a 2D scatter plot represents authors by citations and years.

Pivotpath [131] models a scientific collection as a tripartite graph consisting of three type of nodes: authors, papers, and keywords. It shows relations between them, including citations between papers. The visualization is a node-link diagram with a 2D layout divided into three regions, one for each node type. A pivoting interaction enables the navigation in the tripartite information space. A similar interaction is used also by the above-mentioned (Section 3.4.3) PivotSlice [99], which models a scientific collection as a multivariate network.

3.5.4 Sequential approaches and multiple views

Other examples from the literature we surveyed show a loose coupling of multiple data types. They either apply the same algorithm to different data types of the same document collection, or sequentially apply different analytic algorithms for specific data types and visualize the results once at a time, or juxtapose multiple views with different degrees of coordinated interactivity.

GraphaEL [78], for example, has been demonstrated by applying its evolving graph layout not only to a citation network, but also to a co-authorship network and a topic network. Analogously, White et al. [132] show how SOMs and PFNETs can be applied to both co-citation networks and co-occurrence networks, by switching the similarity metric used by the mapping algorithm. The DIVA system [133] is capable of mapping patents or scientific literature documents to a 2D plane for visual cluster analysis. It features two mapping algorithms (SOM or force-directed layout) and two similarity metrics (citation or co-occurrence). The approach by Ke et al. [134] exploits a node-link diagram to visualize the citation network as well as the co-authorship network, enriched with publication or citation counts. Chou and Yang [135] demonstrate the use of radial node-link diagrams and sunbursts to navigate an ego-centric citation network and a hierarchical classification.

Spangler et al. [136] present a visual search interface for patent retrieval and combine it with a number of interactive views that allow to view and compare documents and their metadata, e.g., in the form of document scatter plots, or trend spark lines.

Chen [137] illustrates the sequential application of algorithms. At first, a latent semantic indexing (LSI) algorithm analyzes the full text and metadata of scientific papers to build a semantic similarity matrix. Then, a PFNET approach is applied for further data reduction. Finally, the identified thematic fields are used to label the clusters of a co-authorship network.

When the number of data types to be considered for the analysis increases, several systems resort to multiple views, each devoted to a subset of data types. Bivteci [138] is an early example of a system that combines authors, citations, and metadata. It features three views: a node-link diagram of citations, a visualization of papers clustered by topics, and a spatialization of papers according to other metadata. The three views are mainly independent from each other, and can be used each at a time to solve a specific task. The Citiviz system [139] features two views: a hyperbolic tree to visualize the hierarchical classification and a scatter plot to visualize papers by date and rank. It allows the user to explore similarities among documents, based on where their topics fit into the classification system, and to assess their relevance and timeliness.

CiteWiz [140] deals with three data types (authors, citations, and metadata) and features three different views. The first view shows authors depicted as human-like glyphs arranged along a time line, scaled according to their impact (i.e. citation count). The second view shows the co-occurrence keyword network as a node-link diagram. The third view is based on Growing Polygons, a radial visualization technique and depicts authors productivity and mutual influence, measured through citations. All views are interactive and can be docked simultaneously within the main interface, but there is no coordination between the views. PaperCube [55] features two main modes, each corresponding to the bibliographic entity to be shown (namely, papers or authors). For each mode, several visual encodings are available, for example time lines, tree maps, node-link diagrams with different layouts, and CircleViews [54] (see also Section 3.2.1). When the user switches between modes and views, the only interactive coordination provided is that filters and highlighted selections are kept, to maintain the analytical focus.

FacetLens [141] is a multi-faceted searching and browsing system that depicts several data aspects with different techniques: mostly coloured circles, but also a containment visualization for hierarchies (e.g., topic classification, or geographic location), and bar charts for temporal aggregates (e.g., number of citations per year). Facets can be interactively manipulated for highlighting, selecting, and filtering attributes and values across different documents. The PatVis system [23] visualizes patent search results and supports query refinement based on eleven views. They span from bar charts and line plots to word clouds, node-link diagrams, tree maps, and geographic maps (Figure 6d). All selection and filtering interactions are coordinated, and a dedicated view shows a graph-based representation of the query as filtering constraints and Boolean operators.

Nazemi et al. [142] propose an adaptive semantic visualization for bibliographic query results. The system is capable of querying a digital library and processing the results to clean data and to compute lightweight semantics. The latter are used to automatically select the most appropriate visualization (one of node-link diagrams, tree maps, or line plots). A visualization cockpit allows the user to further combine several views, coordinated with an enhanced brushing and linking interaction.

3.5.5 *Tight integration*

In this last section, we present integrated approaches that feature a seamless intertwining of several automated algorithms and interactive visualizations, and support the identification of complex patterns involving different data aspects of patents and scientific documents.

The system by Dunne at al. [143] features an integration of visualization techniques and natural language processing algorithms to summarize and cluster documents. It integrates a reference management component and complements the NLP analysis with the computation of citation statistics and network measures, enabling the visualization of citation patterns and clusters. Furthermore, a text visualization provides a summary of incoming citations and outgoing references of the selected document, helping users to make sense of the citation context.

The capability to make sense of relations between documents beyond pure quantitative metrics (e.g., citation counts) is an added value for the visual analysis of patents and scientific literature. Uren et al. [144] propose a sense-making tool to assist the analysis of argument relationships across multiple papers. Their tool does not integrate any automatic NLP algorithms, but provides the user with an interactive visualization to build, annotate, and query maps of concepts and claims, where the citation links can be extended with the type of relation (e.g., similarity, causality) and the polarity (e.g., agrees/disagrees). Schäfer and Spurk [145] propose the application of a sentiment analysis algorithm for classifying citations and drawing a typed citation graph in which citations are coloured according to their polarity. With further integration of text analysis algorithms, it is possible to automatically build a topic influence map. Dietz et al. [146] propose an extended LDA (Latent Dirichlet Analysis) algorithm to identify topics and mutual influences between cited and citing papers. Then, the citation network is visualized as a node-link diagram where less relevant (in the sense of topic influence) links are filtered out. The topic spectrum, i.e., the proportion of topics in each paper, is shown as a colored stacked bar within each node. CitePlag [147] exploits the analysis of the citation context to identify disguised plagiarism techniques, such as paraphrase or cross-language copying. It visualizes the structures of two documents as two stacked bars, and their bibliographic coupling as lines between the positions of in-text references. Moreover, it uses diverse algorithms to identify and visualize particular citation patterns. Görg et al. [30] present a version of Jigsaw for the analysis of scientific literature. Documents are clustered according to two similarity metrics, one based on the analysis of textual content (see Section 3.1), the other one based on connections with entities extracted from metadata (e.g., conferences, affiliations, years). Document clusters are visualized in a node-link diagram, as well as in a grid view, where documents are represented as small squares, ordered and coloured according to their similarity to a selected document. The visualization is enriched with the results of a basic sentiment analysis.

CiteRivers [117] features multiple coordinated views integrated with data mining techniques (Figure 6e). It exploits spectral clustering to hierarchically aggregate documents as well as journals and conference at a granularity level that can be adjusted interactively by the user. The visualization is then enriched by additional computed metrics, such as trendiness (measuring the freshness and the spreading of new ideas), citation entropy (measuring the diversity of a document impact on different research communities), and authors' prolificacy over time.

4 Challenges and Opportunities

Here, we illustrate existing challenges to spur future research, as will new datasets[9]. Some challenges refer to a specific task or data type, while other, more general ones have been derived from the surveyed literature.

Text data: When dealing with large text corpora, NLP and text mining techniques help coping with high dimensionality, identifying and summarizing key topics. Nevertheless, often the most relevant and descriptive terms still need to be visualized as text. Moreover, the visualization needs to also provide semantic context to support correct interpretation. However, providing comprehensive information about documents in a compact way is challenging.

Author data: While scientific documents and patents have unique identifiers (such as the Digital Object Identifier or the unique patent number), author data still suffer from ambiguity because of synonyms (different spelling, same person) and homonyms (same name, different persons). Many approaches disregard this ambiguity, while adequate uncertainty visualization techniques should be utilized to communicate possible ambiguity to users and make them aware of how it can affect the analysis results. Moreover, adhoc visualization techniques might be developed to specifically support name de-duplication and disambiguation.

Citations data: A recent research trend in scientometrics combines citation analysis with sentiment analysis and other text mining techniques, in order to bring the analysis beyond mere citation counts and to consider also the reason beneath a citation (e.g., derivation, confirmation, or confutation). We discussed a couple of preliminary works in Section 3.5.5, but further research is needed to investigate how visualization can help making sense of citations.

Metadata and other data types: Many surveyed techniques lack a specific visualization design for metadata, which are simply treated as generic data. An important example are hierarchical classification schemes. These exist for both patents and scientific literature (see also Section 1.1), and adequate techniques are available for visualizing them [148]. Nevertheless, only few works adopt, refine, or develop techniques for visualizing classification data. Other data types are just ignored in most approaches. Images, for example, are a relevant content of both patents and scientific documents, but often they are not properly considered (with notable exceptions, such as Document Cards [27]).

Connectional tasks: In order to support complex connection discovery, such as finding structural or cause-effect relations between phenomena, visualization techniques need effective means to combine different data aspects and enable their simultaneous exploration. In this context, **visual scalability** is a major challenge (see also [24], [40], for example). As illustrated in Section 3.5, various aggregations and projections as well as coordinated multiple views are applied to tackle large multi-dimensional data sets. However, there is still the need for a systematic analysis of the usefulness and appropriateness of scalable visualization techniques for scientific literature and patent data. Moreover, many of the approaches we surveyed make limited use of **user interaction** to support the analysis task. Some approaches use a loose coupling of views (compare Section 3.5.4) and other approaches present a seamless intertwining of automatic algorithms and interactive visualizations (compare Section 3.5.5). A systematic investigation into the applicability of various interaction techniques (Detail-On-Demand, Focus+Context, Coordinated Multiple Views) to solve particular problems in this domain would open new research possibilities. As illustrated in Section 3.5.5, several approaches follow the Visual Analytics paradigm. However, automatic mining algorithms should not be treated as a black box, but allow the user to interact with the different parameter settings [149]. Furthermore, such an approach should support the various steps in the knowledge generation process [150]. Consequently, there is open space for future investigation in the field of patents and scientific literature. In particular, many surveyed approaches support synoptic tasks involving temporal patters and they also address the problem of understanding the connections between non-temporal and temporal dynamics across multiple data types. Nonetheless, interactive visualization and automatic analysis need to be integrated further in order to enable users to formulate and test temporal hypotheses, as well as to build or select models of time-oriented data. The pursuit of this research direction would enable not only the early identification of emerging technologies and research fronts, but also their **prediction**.

General challenges: Quantitative and qualitative **evaluations** are important to gauge the effectiveness of new developments in science. Nevertheless, only a few works, among those we surveyed, include an in-depth evaluation of the proposed technique by user studies. Hence, more empirical research is needed to determine which techniques work better for which data types, analysis tasks, users, and combinations thereof. Visualizing patents and scientific literature is definitely a **multidisciplinary** problem. On the one hand, the visualization community is interested to contribute with innovative visual and Visual Analytics approaches. On the other hand, domain experts are eager to explore the various data bases or resources according to different criteria. Multidisciplinary approaches need a common understanding of the various disciplines, their aims, and their tasks. Therefore, a systematic approach, like the data-users-tasks design triangle [151], could be useful to guide and steer the different experts involved.

5 Conclusion

In this survey, we investigated different interactive visualization approaches of patents and scientific articles, ranging from explanation tools to sophisticated exploration methods. We categorized the survey according to two aspects: (a) data types (text, citations, authors, metadata, and combinations thereof) and (b) analyses types (exploring and

9. http://vispubdata.org

comparing single entities, investigating relations between entities, finding patterns on various levels, exploring the temporal dynamics, and finding complex connections between phenomena). Obviously, our aspects and categories are not mutually exclusive and some approaches fit to several aspects. To solve this issue, we identified the main aspect of an approach, discussed it in that section, and referred to it in the others.

The visual representations are obviously and self-evidently dependent on the data types and the analyses tasks. Graph and graph-based visualization (like node-link diagrams), matrix-based visualizations as well as scatter plots and scatter plot matrices are applied to explore the relations of various entities and citations as well as to detect patterns. Different derivatives of streamgraphs and line charts are used to analyze temporal dynamics. Furthermore, animation and small multiples are used instead of simple time lines. To tackle the complexity of patents and scientific articles, various kinds of glyphs as well as maps and landscapes are proposed. In case multiple variables as well as multiple data types are explored, aggregations and projections as well as multiple and/or sequential views with and without tight integration and interactions are proposed. The text analysis methods range from manual and automatic annotations to sophisticated NLP techniques. However, visual means to fine-tune the parameter settings are seldom applied.

Acknowledgments

Parts of this work were funded by the German Science Foundation (DFG) through the Priority Program "Scalable Visual Analytics" and by the Austrian Research Promotion Agency (FFG, grant no. 835937).

References

[1] D. Bonino, A. Ciaramella, and F. Corno, "Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics," *World Patent Inform.*, vol. 32, no. 1, pp. 30 – 38, 2010.

[2] A. Abbas, L. Zhang, and S. U. Khan, "A literature review on the state-of-the-art in patent analysis," *World Patent Inform.*, vol. 37, pp. 3 – 13, 2014.

[3] Y. Yang, L. Akers, T. Klose, and C. B. Yang, "Text mining and visualization tools - Impressions of emerging capabilities," *World Patent Inform.*, vol. 30, no. 4, pp. 280 – 293, 2008.

[4] M. Lupu, K. Mayer, J. Tait, and A. J. Trippe, Eds., *Current Challenges in Patent Information Retrieval*. Springer, 2011.

[5] C. Chen, *Mapping Scientific Frontiers - The Quest for Knowledge Visualization, 2nd Edition*. Springer, 2013.

[6] K. Börner, *Atlas of Science: Visualizing What We Know*. The MIT Press, 2010.

[7] M. A. Hearst, *Search User Interfaces*, 1st ed. New York, NY, USA: Cambridge University Press, 2009.

[8] A. B. Alencar, M. C. F. de Oliveira, and F. V. Paulovich, "Seeing beyond reading: a survey on visual text analytics," *WIREs Data Mining and Knowledge Discovery*, vol. 2, no. 6, pp. 476–492, 2012.

[9] K. Kucher and A. Kerren, "Text visualization techniques: Taxonomy, visual survey, and community insights," in *Proc. IEEE Pacific Visualization Symp. (PacificVis)*, 2015, pp. 117–121.

[10] B. Shneiderman, "The eyes have it: a task by data type taxonomy for information visualizations," in *Proc. IEEE Symp. Visual Languages*, 1996, pp. 336–343.

[11] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the web," Stanford University InfoLab, Tech. Rep. 66, 1999.

[12] J. E. Hirsch, "An index to quantify an individual's scientific research output," *PNAS*, vol. 102, no. 46, pp. 16 569–16 572, 2005.

[13] N. Andrienko and G. Andrienko, *Exploratory analysis of spatial and temporal data: a systematic approach*. Berlin: Springer, 2006.

[14] C. D. Manning, P. Raghavan, H. Schütze *et al.*, *Introduction to information retrieval*. Cambridge University Press, 2008, vol. 1, no. 1.

[15] C. D. Manning and H. Schütze, *Foundations of statistical natural language processing*. MIT Press, 1999, vol. 1.

[16] J. H. Martin and D. Jurafsky, *Speech and Language Processing. An introduction to natural language processing, computational linguistics, and speech recognition*. Pearson Education Limited, 2014, vol. 2.

[17] S. Havre, E. Hetzler, K. Perrine, E. Jurrus, and N. Miller, "Interactive visualization of multiple query results," in *Proc. IEEE Symp. Inform. Visualization (InfoVis)*. IEEE, 2001, pp. 105–112.

[18] J. Chuang, D. Ramage, C. Manning, and J. Heer, "Interpretation and trust: Designing model-driven visualizations for text analysis," in *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI)*. New York, NY, USA: ACM, 2012, pp. 443–452.

[19] J. Choo, C. Lee, C. K. Reddy, and H. Park, "UTOPIAN: User-driven topic modeling based on interactive nonnegative matrix factorization," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 12, pp. 1992–2001, 2013.

[20] W. Dou, L. Yu, X. Wang, Z. Ma, and W. Ribarsky, "HierarchicalTopics: Visually exploring large text collections using topic hierarchies," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 12, pp. 2002–2011, 2013.

[21] M. A. Hearst, "TileBars: visualization of term distribution information in full text information access," in *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI)*. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1995, pp. 59–66.

[22] L. T. Nowell, R. K. France, D. Hix, L. S. Heath, and E. A. Fox, "Visualizing search results: some alternatives to query-document similarity," in *Proc. 19th Annu. Int. ACM SIGIR Conf. Research and Develop. in Inform. Retrieval*. New York, NY, USA: ACM, 1996, pp. 67–75.

[23] S. Koch, H. Bosch, M. Giereth, and T. Ertl, "Iterative integration of visual insights during scalable patent search and analysis," *IEEE Trans. Vis. Comput. Graphics*, vol. 17, no. 5, pp. 557–569, 2011.

[24] K. A. Olsen, R. R. Korfhage, K. M. Sochats, M. B. Spring, and J. G. Williams, "Visualization of a document collection: The VIBE system," *Inform. Process. & Manage.*, vol. 29, no. 1, pp. 69–81, 1993.

[25] R. R. Korfhage, "To see, or not to see – Is That the query?" in *Proc. 14th Annu. Int. ACM SIGIR Conf. Research and Develop. in Inform. Retrieval*. New York, NY, USA: ACM, 1991, pp. 134–141.

[26] G. Costagliola and V. Fuccella, "CyBiS: A novel interface for searching scientific documents," in *15th Int. Conf. Inform. Visualisation (IV)*, 2011, pp. 276–281.

[27] H. Strobelt, D. Oelke, C. Rohrdantz, A. Stoffel, D. A. Keim, and O. Deussen, "Document cards: A top trumps visualization for documents," *IEEE Trans. Vis. Comput. Graphics*, vol. 15, no. 6, pp. 1145–1152, 2009.

[28] R. Nakazawa, T. Itoh, and T. Saito, "A visualization of research papers based on the topics and citation network," in *Proc. 19th Conf. Inform. Visualization (IV)*, 2015.

[29] C. Görg, H. Tipney, K. Verspoor, W. A. Baumgartner Jr, K. B. Cohen, J. Stasko, and L. E. Hunter, "Visualization and language processing for supporting analysis across the biomedical literature," in *Knowledge-Based and Intelligent Information and Engineering Systems*. Springer, 2010, pp. 420–429.

[30] C. Görg, Z. Liu, J. Kihm, J. Choo, H. Park, and J. Stasko, "Combining computational analyses and interactive visualization for document exploration and sensemaking in Jigsaw," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 10, pp. 1646–1663, 2013.

[31] P. Riehmann, M. Potthast, B. Stein, and B. Froehlich, "Visual assessment of alleged plagiarism cases," *Computer Graphics Forum*, 2015.

[32] F. Osborne, E. Motta, and P. Mulholland, "Exploring scholarly data with Rexplore," in *Int. Semantic Web Conf. (ISWC)*. Springer, 2013, pp. 460–477.

[33] B. Lee, M. Czerwinski, G. Robertson, and B. B. Bederson, "Understanding research trends in conferences using PaperLens," in *Extended Abstracts on Human Factors in Computing Systems (CHI EA)*. New York, NY, USA: ACM, 2005, pp. 1969–1972.

[34] G. S. Davidson, B. Hendrickson, D. K. Johnson, C. E. Meyers, and B. N. Wylie, "Knowledge mining with VxInsight: Discovery

through interaction," *J. Intell. Inform. Syst.*, vol. 11, no. 3, pp. 259–285, 1998.

[35] P. C. Wong, B. Hetzler, C. Posse, M. Whiting, S. Havre, N. Cramer, A. Shah, M. Singhal, A. Turner, and J. Thomas, "IN-SPIRE InfoVis 2004 contest entry," in *Proc. IEEE Symp. Inform. Visualization (InfoVis)*. IEEE, 2004.

[36] P. DeCamp, A. Frid-Jimenez, J. Guiness, and D. Roy, "Gist icons: Seeing meaning in large bodies of literature," in *Poster Proc. IEEE Symp. Inform. Visualization (InfoVis)*, 2005.

[37] Y. Wu, T. Provan, F. Wei, S. Liu, and K.-L. Ma, "Semantic-preserving word clouds by seam carving," *Computer Graphics Forum*, vol. 30, no. 3, pp. 741–750, 2011.

[38] J. Chuang, C. D. Manning, and J. Heer, "Termite: Visualization techniques for assessing textual topic models," in *Proc. Int. Working Conf. Advanced Visual Interfaces (AVI)*. ACM, 2012, pp. 74–77.

[39] J. Chuang, D. Ramage, D. A. McFarland, C. D. Manning, and J. Heer, "Large-scale examination of academic publications using statistical models," in *Proc. of the AVI Workshop on Supporting Asynchronous Collaboration in Visual Analytics Systems*, 2012.

[40] W. Dou, X. Wang, R. Chang, and W. Ribarsky, "ParallelTopics: A probabilistic approach to exploring document collections," in *Proc. IEEE Conf. Visual Analytics Science and Technology (VAST)*. IEEE, 2011, pp. 231–240.

[41] X. Jiang and J. Zhang, "A text visualization method for cross-domain research topic mining," *J. of Visualization*, pp. 1–16, 2016.

[42] B. Gretarsson, J. O'Donovan, S. Bostandjiev, T. Höllerer, A. Asuncion, D. Newman, and P. Smyth, "TopicNets: Visual analysis of large text corpora with topic modeling," *ACM Trans. Intell. Syst. Technol.*, vol. 3, no. 2, pp. 23:1–23:26, Feb. 2012.

[43] H. Lee, J. Kihm, J. Choo, J. Stasko, and H. Park, "iVisClustering: An interactive visual document clustering via topic modeling," *Computer Graphics Forum*, 2012.

[44] P. Oesterling, G. Scheuermann, S. Teresniak, G. Heyer, S. Koch, T. Ertl, and G. H. Weber, "Two-stage framework for a topology-based projection and visualization of classified document collections," in *Proc. IEEE Symp. Visual Analytics Science and Technology (VAST)*. IEEE, 2010, pp. 91–98.

[45] D. Oelke, H. Strobelt, C. Rohrdantz, I. Gurevych, and O. Deussen, "Comparative exploration of document collections: a visual analytics approach," *Computer Graphics Forum*, vol. 33, no. 3, pp. 201–210, 2014.

[46] T. Kohonen, S. Kaski, K. Lagus, J. Salojarvi, J. Honkela, V. Paatero, and A. Saarela, "Self organization of a massive document collection," *IEEE Trans. Neural Netw.*, vol. 11, no. 3, pp. 574–585, 2000.

[47] D. Fried and S. G. Kobourov, "Maps of computer science," in *IEEE Pacific Visualization Symp. (PacificVis)*, 2014, pp. 113–120.

[48] A. Skupin, "A cartographic approach to visualizing conference abstracts," *IEEE Comput. Graph. Appl.*, vol. 22, no. 1, pp. 50–58, 2002.

[49] A. Skupin, "The world of geography: Visualizing a knowledge domain with cartographic means," *PNAS*, vol. 101, no. Suppl 1, pp. 5274–5278, 2004.

[50] K. K. Mane and K. Börner, "Mapping topics and topic bursts in PNAS," *PNAS*, vol. 101, no. Suppl 1, pp. 5287–5290, 2004.

[51] S. Havre, B. Hetzler, and L. Nowell, "ThemeRiver: visualizing theme changes over time," in *Proc. IEEE Symp. Inform. Visualization (InfoVis)*, 2000, pp. 115–123.

[52] A. Ahmed, T. Dwyer, C. Murray, L. Song, and Y. X. Wu, "InfoVis 2004 Contest: Wilmascope graph visualisation," in *Proc. IEEE Symp. Inform. Visualization (InfoVis)*. IEEE, 2004.

[53] S. Ankam, W. Dou, D. Strumsky, D. X. Wang, T. Rabinowitz, and W. Zadrozny, "Exploring emerging technologies using patent data and patent classification," in *Proc. IEEE VIS Workshop on Interactive Visual Text Analytics*, 2012.

[54] P. Bergström and E. J. Whitehead Jr, "CircleView: scalable visualization and navigation of citation networks," in *Proc. Symp. Interactive Visual Inform. Collections and Activity IVICA*, 2006.

[55] P. Bergström and D. Atkinson, "Augmenting the exploration of digital libraries with web-based visualizations," in *Proc. 4th Int. Conf. Digital Inform. Manage. (ICDIM)*, 2009, pp. 1–7.

[56] N. J. van Eck and L. Waltman, "CitNetExplorer: A new software tool for analyzing and visualizing citation networks," *J. Informetrics*, vol. 8, no. 4, pp. 802 – 823, 2014.

[57] K. W. Boyack, B. N. Wylie, G. S. Davidson, and D. K. Johnson, "Analysis of patent databases using VxInsight TM," in *New Paradigms in Inform. Visualization and Manipulation, a Workshop at the 9th Int. Conf. on Inform. and Knowledge Manage. (CIKM)*, 2000.

[58] M. Rosvall and C. T. Bergstrom, "Mapping change in large networks," *PLoS ONE*, vol. 5, no. 1, p. e8694, 01 2010.

[59] E. Yan and Y. Ding, "Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and coword networks relate to each other," *J. Amer. Soc. Inform. Sci. and Technology*, vol. 63, no. 7, pp. 1313–1326, 2012.

[60] J. D. Mackinlay, R. Rao, and S. K. Card, "An organic user interface for searching citation links," in *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI)*. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1995, pp. 67–73.

[61] E. Garfield, "Historiographic mapping of knowledge domains literature," *J. Inform. Sci.*, vol. 30, no. 2, pp. 119–145, 2004.

[62] N. Henry, H. Goodell, N. Elmqvist, and J.-D. Fekete, "20 years of four HCI conferences: A visual exploration," *Int. J. Human-Computer Interaction*, vol. 23, no. 3, pp. 239–285, 2007.

[63] A. Aris, B. Shneiderman, V. Qazvinian, and D. Radev, "Visual overviews for discovering key papers and influences across research fronts," *J. Amer. Soc. Inform. Sci. and Technology*, vol. 60, no. 11, pp. 2219–2228, 2009.

[64] L. Marks, J. A. Hussell, T. M. McMahon, and R. E. Luce, "ActiveGraph: A digital library visualization tool," *Int. J. Digital Libraries*, vol. 5, no. 1, pp. 57–69, 2005.

[65] H. Small, "Visualizing science by citation mapping," *J. Amer. Soc. Inform. Sci.*, vol. 50, no. 9, pp. 799–813, 1999.

[66] T.-T. Chen and L. C. Hsieh, "On visualization of cocitation networks," in *11th Int. Conf. Inform. Visualization (IV)*, 2007, pp. 470–475.

[67] M. Delest, T. Munzner, D. Auber, and J.-P. Domenger, "Exploring InfoVis publication history with Tulip," in *Proc. IEEE Symp. Inform. Visualization (InfoVis)*. IEEE, 2004.

[68] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *PNAS*, vol. 105, no. 4, pp. 1118–1123, 2008.

[69] L. Shi, H. Tong, J. Tang, and C. Lin, "VEGAS: Visual influence graph summarization on citation networks," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 12, pp. 3417–3431, 2015.

[70] C. Chen and S. Morris, "Visualizing evolving networks: Minimum spanning trees versus pathfinder networks," in *Proc. IEEE Symp. Inform. Visualization (InfoVis)*. IEEE, 2003, pp. 67–74.

[71] S. Noel, C.-H. H. Chu, and V. Raghavan, "Co-citation count vs correlation for influence network visualization," *Inform. Visualization*, vol. 2, no. 3, pp. 160–170, 2003.

[72] J. Zhang, C. Chen, and J. Li, "Visualizing the intellectual structure with paper-reference matrices," *IEEE Trans. Vis. Comput. Graphics*, vol. 15, no. 6, pp. 1153–1160, 2009.

[73] U. Brandes and T. Willhalm, "Visualization of bibliographic networks with a reshaped landscape metaphor," in *Proc. Symp. Data Visualisation (VISSYM)*. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2002, pp. 159–ff.

[74] N. J. van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523–538, 2009.

[75] C. Chen, "Searching for intellectual turning points: Progressive knowledge domain visualization," *PNAS*, vol. 101, no. Suppl 1, pp. 5303–5310, 2004.

[76] B. W. Herr, R. J. Duhon, K. Börner, E. F. Hardy, and S. Penumarthy, "113 years of physical review: Using flow maps to show temporal and topical citation patterns," in *Proc. 12th Int. Conf. Inform. Visualisation (IV)*. IEEE, 2008, pp. 421–426.

[77] J. Matejka, T. Grossman, and G. Fitzmaurice, "Citeology: Visualizing paper genealogy," in *Extended Abstracts on Human Factors in Computing Systems (CHI EA)*. New York, NY, USA: ACM, 2012, pp. 181–190.

[78] C. Erten, P. J. Harding, S. G. Kobourov, K. Wampler, and G. Yee, "GraphAEL: Graph animations with evolving layouts," in *Graph Drawing*, ser. LNCS, G. Liotta, Ed. Springer Berlin Heidelberg, 2004, vol. 2912, pp. 98–110.

[79] C. Chen and J. Kuljis, "The rising landscape: A visual exploration of superstring revolutions in physics," *J. Amer. Soc. Inform. Sci. and Technology*, vol. 54, pp. 435–446, 2003.

[80] H. G. Small, "A co-citation model of a scientific specialty: A longitudinal study of collagen research," *Social studies of science*, vol. 1, pp. 139–166, 1977.

[81] N. Shibata, Y. Kajikawa, Y. Takeda, and K. Matsushima, "Detecting emerging research fronts based on topological measures in

citation networks of scientific publications," *Technovation*, vol. 28, no. 11, pp. 758 – 775, 2008.

[82] J. Abello, S. Hadlak, H. Schumann, and H.-J. Schulz, "A modular degree-of-interest specification for the visual analysis of large dynamic networks," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 3, pp. 337–350, March 2014.

[83] C. Chen, J. Zhang, W. Zhu, and M. Vogeley, "Delineating the citation impact of scientific discoveries," in *Proc. 7th ACM/IEEE-CS joint Conf. Digital Libraries (DL)*. ACM, 2007, pp. 19–28.

[84] C. Chen, "CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature," *J. Amer. Soc. Inform. Sci. and Technology*, vol. 57, no. 3, pp. 359–377, 2006.

[85] H. Kang, C. Plaisant, B. Lee, and B. B. Bederson, "NetLens: Iterative exploration of content-actor network data," *Inform. Visualization*, vol. 6, no. 1, pp. 18–31, 2007.

[86] Z. Chinchilla-Rodríguez, B. Vargas-Quesada, Y. Hassan-Montero, A. González-Molina, and F. Moya-Anegóna, "New approach to the visualization of international scientific collaboration," *Inform. Visualization*, vol. 9, no. 4, pp. 277–287, 2010.

[87] N. Henry, J.-D. Fekete, and M. McGuffin, "NodeTrix: a hybrid visualization of social networks," *IEEE Trans. Vis. Comput. Graphics*, vol. 13, no. 6, pp. 1302–1309, 2007.

[88] T.-H. Huang and M. L. Huang, "Analysis and visualization of co-authorship networks for understanding academic collaboration and knowledge domain of individual researchers," in *Proc. Int. Conf. Computer Graphics, Imaging and Visualisation*, 2006, pp. 18–23.

[89] K. Misue, "Visual analysis tool for bipartite networks," in *Knowledge-Based Intell. Inform. and Eng. Syst.*, ser. LNCS. Springer Berlin Heidelberg, 2008, vol. 5178, pp. 871–878.

[90] R. Ichise, H. Takeda, and K. Ueyama, "Community mining tool using bibliography data," in *Proc. 9th Int. Conf. Inform. Visualisation (IV)*. IEEE, 2005, pp. 953–958.

[91] D. Kutz, "Examining the evolution and distribution of patent classifications," in *Proc. 8th Int. Conf. Inform. Visualisation (IV)*, 2004, pp. 983–988.

[92] D. A. Keim, H. Barro, C. Panse, J. Schneidewind, and M. Sips, "Exploring and visualizing the history of InfoVis," in *InfoVis*, 2004.

[93] L. Shi, C. Wang, Z. Wen, H. Qu, C. Lin, and Q. Liao, "1.5d egocentric dynamic network visualization," *IEEE Trans. Vis. Comput. Graphics*, vol. 21, no. 5, pp. 624–637, May 2015.

[94] B. Bach, E. Pietriga, and J.-D. Fekete, "GraphDiaries: Animated transitions and temporal navigation for dynamic networks," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 5, pp. 740–754, May 2014.

[95] J. Heer and A. Perer, "Orion: A system for modeling, transformation and visualization of multidimensional heterogeneous networks," in *Proc. IEEE Conf. Visual Analytics Science and Technology (VAST)*. IEEE, 2011, pp. 51–60.

[96] T. Kurosawa and Y. Takama, "Predicting researchers' future activities using visualization system for co-authorship networks," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intelligence and Intell. Agent Technology (WI-IAT)*. Washington, DC, USA: IEEE Comp. Soc., 2011, pp. 332–339.

[97] F. Beck, S. Koch, and D. Weiskopf, "Visual analysis and dissemination of scientific literature collections with SurVis," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 01, January 2016.

[98] B. Alsallakh, W. Aigner, S. Miksch, and H. Hauser, "Radial Sets: Interactive visual analysis of large overlapping sets," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 12, pp. 2496–2505, 2013.

[99] J. Zhao, C. Collins, F. Chevalier, and R. Balakrishnan, "Interactive exploration of implicit and explicit relations in faceted datasets," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 12, pp. 2080–2089, 2013.

[100] C. Collins, G. Penn, and S. Carpendale, "Bubble sets: Revealing set relations with isocontours over existing visualizations," *IEEE Trans. Vis. Comput. Graphics*, vol. 15, no. 6, pp. 1009–1016, 2009.

[101] M. Giereth, M. Wörner, H. Bosch, P. Baier, and T. Ertl, "Utilization of semantic annotations in interactive user interfaces for large documents," in *GI Jahrestagung (2)*, 2008, pp. 706–711.

[102] M. Hascoët and P. Dragicevic, "Interactive graph matching and visual comparison of graphs and clustered graphs," in *Proc. Int. Working Conf. Advanced Visual Interfaces (AVI)*. ACM, 2012, pp. 522–529.

[103] H. Guo, S. R. Gomez, M. J. Schnitzer, and D. H. Laidlaw, "Visualization to facilitate structured exploration of published findings in rat brain connectivity," in *Poster Proc. IEEE Visualization Conf.*, 2013.

[104] K. V. Nesbitt, "Getting to more abstract places using the metro map metaphor," in *Proc. 8th Int. Conf. Inform. Visualisation (IV)*. IEEE, 2004, pp. 488–493.

[105] B. Wong, S. Choudhury, C. Rooney, R. Chen, and K. Xu, "INVISQUE: Technology and methodologies for interactive information visualization and analytics in large library collections," in *Research and Advanced Technology for Digital Libraries*, ser. LNCS, S. Gradmann, F. Borri, C. Meghini, and H. Schuldt, Eds. Springer Berlin Heidelberg, 2011, vol. 6966, pp. 227–235.

[106] K. Wittenburg, A. Malizia, L. Lupo, and G. Pekhteryev, "Visualizing set-valued attributes in parallel with equal-height histograms," in *Proc. Int. Working Conf. Advanced Visual Interfaces (AVI)*. New York, NY, USA: ACM, 2012, pp. 632–635.

[107] K. Wittenburg and G. Pekhteryev, "Multi-dimensional comparative visualization for patent landscaping," in *IEEE VIS Workshop BusinessVis*, 2015.

[108] B. Shneiderman, D. Feldman, A. Rose, and X. F. Grau, "Visualizing digital library search results with categorical and hierarchical axes," in *Proc. 5th ACM Conf. Digital Libraries (DL)*. New York, NY, USA: ACM, 2000, pp. 57–66.

[109] A. Sallaberry, N. Pecheur, S. Bringay, M. Roche, and M. Teisseire, "Discovering novelty in gene data: from sequential patterns to visualization," in *Advances in Visual Computing*. Springer, 2010, pp. 534–543.

[110] M. Giereth, H. Bosch, and T. Ertl, "A 3D treemap approach for analyzing the classificatory distribution in patent portfolios," in *Proc. IEEE Symp. Visual Analytics Science and Technology (VAST)*, 2008, pp. 189–190.

[111] K. Börner, E. Hardy, B. Herr, T. Holloway, and W. B. Paley, "Taxonomy visualization in support of the semi-automatic validation and optimization of organizational schemas," *J. Informetrics*, vol. 1, no. 3, pp. 214–225, 2007.

[112] A. Perer, I. Guy, E. Uziel, I. Ronen, and M. Jacovi, "Visual social network analytics for relationship discovery in the enterprise," in *Proc. IEEE Conf. Visual Analytics Science and Technology (VAST)*. IEEE, 2011, pp. 71–79.

[113] S.-H. Chen, M.-H. Huang, and D.-Z. Chen, "Visualization of the technology evolution in smart grid," in *Proc. Technology Manage. in the Energy Smart World (PICMET)*, 2011, pp. 1–7.

[114] C. Chen and R. Paul, "Visualizing a knowledge domain's intellectual structure," *Computer*, vol. 34, no. 3, pp. 65–71, 2001.

[115] H. Sharara, A. Sopan, G. Namata, L. Getoor, and L. Singh, "G-PARE: a visual analytic tool for comparative analysis of uncertain graphs," in *Proc. IEEE Conf. Visual Analytics Science and Technology (VAST)*. IEEE, 2011, pp. 61–70.

[116] Z. Shen, M. Ogawa, S. T. Teoh, and K.-L. Ma, "BiblioViz: A system for visualizing bibliography information," in *Proc. Asia-Pacific Symp. Inform. Visualisation (APVis)*. Darlinghurst, Australia, Australia: Australian Comp. Soc., Inc., 2006, pp. 93–102.

[117] F. Heimerl, Q. Han, S. Koch, and T. Ertl, "CiteRivers: Visual analysis of citation patterns," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 01, January 2016.

[118] B. Alper, N. Riche, G. Ramos, and M. Czerwinski, "Design study of LineSets, a novel set visualization technique," *IEEE Trans. Vis. Comput. Graphics*, vol. 17, no. 12, pp. 2259–2267, Dec 2011.

[119] H. D. White and K. W. McCain, "Visualizing a discipline: An author co-citation analysis of information science, 1972-1995," *J. Amer. Soc. Inform. Sci.*, vol. 49, no. 4, pp. 327–355, 1998.

[120] X. Lin, H. D. White, and J. Buzydlowski, "Real-time author co-citation mapping for online searching," *Inform. process. & manage.*, vol. 39, no. 5, pp. 689–706, 2003.

[121] X. Li, H. Chen, Z. Huang, and M. Roco, "Patent citation network in nanotechnology (1976 - 2004)," *J. Nanoparticle Research*, vol. 9, no. 3, pp. 337–352, 2007.

[122] F. Windhager, A. Amor-Amoros, M. Smuc, P. Federico, L. Zenk, and S. Miksch, "A concept for the exploratory visualization of patent network dynamics," in *Proc. Int. Conf. Information Visualization Theory Appl. (IVAPP)*, 2015, pp. 268–273.

[123] T. Nagel, E. Duval, and A. Vande Moere, "Interactive exploration of geospatial network visualization," in *Extended Abstracts on Human Factors in Computing Systems (CHI EA)*. ACM, 2012, pp. 557–572.

[124] M. Rosvall and C. T. Bergstrom, "Multilevel compression of random walks on networks reveals hierarchical organization in

large integrated systems," *PLoS ONE*, vol. 6, no. 4, p. e18209, 04 2011.

[125] T. Honkela, J. Laaksonen, H. Törrö, and J. Tenhunen, "Media map: A multilingual document map with a design interface," in *Advances in Self-Organizing Maps*. Springer, 2011, pp. 247–256.

[126] I. Jusufi, A. Kerren, J. Liu, and B. Zimmer, "Visual exploration of relationships between document clusters," in *Proc. Int. Conf. Information Visualization Theory Appl. (IVAPP)*, 2014, pp. 195–203.

[127] S. A. Morris, G. Yen, Z. Wu, and B. Asnake, "Time line visualization of research fronts," *J. Amer. Soc. Inform. Sci. and Technology*, vol. 54, no. 5, pp. 413–422, 2003.

[128] T. Sun and S. A. Morris, "Timeline and crossmap visualization of patents," in *Proc. 4th Int. Conf. Webometrics, Informetrics and Scientometrics (WIS)*, 2008.

[129] F. van Ham, "Case study: Visualizing visualization," in *Proc. IEEE Symp. Inform. Visualization (InfoVis)*. IEEE, 2004.

[130] J. Tyman, G. P. Gruetzmacher, and J. Stasko, "InfoVisExplorer," in *Proc. IEEE Symp. Inform. Visualization (InfoVis)*. IEEE, 2004.

[131] M. Dörk, N. Henry Riche, G. Ramos, and S. Dumais, "PivotPaths: Strolling through faceted information spaces," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 12, pp. 2709–2718, 2012.

[132] H. D. White, X. Lin, J. W. Buzydlowski, and C. Chen, "User-controlled mapping of significant literatures," *PNAS*, vol. 101, no. Suppl 1, pp. 5297–5302, 2004.

[133] S. Morris, C. DeYong, Z. Wu, S. Salman, and D. Yemenu, "DIVA: a visualization system for exploring document databases for technology forecasting," *Comput. & Ind. Eng.*, vol. 43, no. 4, pp. 841 – 862, 2002.

[134] W. Ke, K. Borner, and L. Viswanath, "Major information visualization authors, papers and topics in the ACM library," in *Proc. IEEE Symp. Inform. Visualization (InfoVis)*. IEEE, 2004.

[135] J. K. Chou and C. K. Yang, "PaperVis: Literature review made easy," *Computer Graphics Forum*, vol. 30, no. 3, pp. 721–730, 2011.

[136] S. Spangler, Y. Chen, J. Kreulen, S. Boyer, T. Griffin, A. Alba, L. Kato, A. Lelescu, and S. Yan, "SIMPLE: Interactive analytics on patent data," in *Proc. IEEE Int. Conf. Data Mining Workshops*, 2010, pp. 426–433.

[137] C. Chen, "Visualising semantic spaces and author co-citation networks in digital libraries," *Inform. process. & manage.*, vol. 35, no. 3, pp. 401–420, 1999.

[138] D. Modjeska, V. Tzerpos, P. Faloutsos, and M. Faloutsos, "BIVTECI: A bibliographic visualization tool," in *Proc. Conf. Centre for Advanced Studies on Collaborative Research (CASCON)*. IBM Press, 1996.

[139] N. Kampanya, R. Shen, S. Kim, C. North, and E. A. Fox, "Citiviz: A visual user interface to the CITIDEL system," in *Research and Advanced Technology for Digital Libraries*, ser. LNCS, R. Heery and L. Lyon, Eds. Springer Berlin Heidelberg, 2004, vol. 3232, pp. 122–133.

[140] N. Elmqvist and P. Tsigas, "CiteWiz: A tool for the visualization of scientific citation networks," *Inform. Visualization*, vol. 6, no. 3, pp. 215–232, 2007.

[141] B. Lee, G. Smith, G. G. Robertson, M. Czerwinski, and D. S. Tan, "FacetLens: exposing trends and relationships to support sensemaking within faceted datasets," in *Proc. SIGCHI Conf. Human Factors in Computing Systems*. ACM, 2009, pp. 1293–1302.

[142] K. Nazemi, R. Retz, J. Bernard, J. Kohlhammer, and D. Fellner, "Adaptive semantic visualization for bibliographic entries," in *Advances in Visual Computing*, ser. LNCS, G. Bebis, R. Boyle, B. Parvin, D. Koracin, B. Li, F. Porikli, V. Zordan, J. Klosowski, S. Coquillart, X. Luo, M. Chen, and D. Gotz, Eds. Springer Berlin Heidelberg, 2013, vol. 8034, pp. 13–24.

[143] C. Dunne, B. Shneiderman, R. Gove, J. Klavans, and B. Dorr, "Rapid understanding of scientific paper collections: integrating statistics, text analytics, and visualization," *J. Amer. Soc. Inform. Sci. and Technology*, 2012.

[144] V. Uren, S. B. Shum, M. Bachler, and G. Li, "Sensemaking tools for understanding research literatures: Design, implementation and user evaluation," *Int. J. Human-Computer Studies*, vol. 64, no. 5, pp. 420 – 445, 2006.

[145] U. Schäfer and C. Spurk, "TAKE scientist's workbench: Semantic search and citation-based visual navigation in scholar papers," in *Proc. IEEE 4th Int. Conf. Semantic Computing (ICSC)*, 2010, pp. 317–324.

[146] L. Dietz, S. Bickel, and T. Scheffer, "Unsupervised prediction of citation influences," in *Proc. 24th Int. Conf. Mach. Learning (ICML)*. ACM, 2007, pp. 233–240.

[147] B. Gipp, N. Meuschke, C. Breitinger, J. Pitman, and A. Nürnberger, "Web-based demonstration of semantic similarity detection using citation pattern visualization for a cross language plagiarism case," in *Proc. 16th Int. Conf. Enterprise Inform. Syst. (ICEIS)*, 2014, pp. 677–683.

[148] H. J. Schulz, "Treevis.net: A tree visualization reference," *IEEE Comput. Graph. Appl.*, vol. 31, no. 6, pp. 11–15, Nov 2011.

[149] T. Mühlbacher, H. Piringer, S. Gratzl, M. Sedlmair, and M. Streit, "Opening the black box: Strategies for increased user involvement in existing algorithm implementations," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 12, pp. 1643–1652, 2014.

[150] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim, "Knowledge generation model for visual analytics," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 12, pp. 1604 – 1613, Dec. 2014.

[151] S. Miksch and W. Aigner, "A matter of time: Applying a data-users-tasks design triangle to visual analytics of time-oriented data," *Computers & Graphics*, vol. 38, pp. 286–290, 2014.

Paolo Federico is a doctoral student in the Institute of Software Technology and Interactive Systems at Vienna University of Technology, Austria. His main research interests are information visualization and visual analytics, in particular of time-oriented and network data. Federico holds a Masters in Telecommunications Engineering from the "Federico II" University of Naples, Italy.

Florian Heimerl is a doctoral student in the Institute for Visualization and Interactive Systems at the University of Stuttgart, Germany. His research focuses on visual analytics and visual text analysis. Heimerl has a Diplom degree in computational linguistics from the University of Stuttgart.

Steffen Koch is a research associate at the Institute for Visualization and Interactive Systems, University of Stuttgart, Germany, where he received his doctorate in computer science. His research interests comprise visualization in general, with foci on visual analytics for text/documents, visualization in the digital humanities, and interactive visualization support for data mining/machine learning.

Silvia Miksch is University Professor at Institute of Software Technology and Interactive Systems, Vienna University of Technology. She served as paper co-chair of several conferences including IEEE VAST 2010 & 2011 and EuroVis 2012, and on the editorial board of several journals including IEEE TVCG. Her main research interests are Visualization/Visual Analytics (in particular Focus+Context and Interaction methods) and Time.