
Automatically Generating Citation Graphs (and Variants) for Systematic Reviews

Sven Groppe, Lina Hartung

Institute of Information Systems (IFIS), University of Lübeck, Ratzeburger Allee 160, D-23562 Lübeck, Germany,
groppe@ifis.uni-luebeck.de, lina.hartung@protonmail.com

ABSTRACT

Citation graphs visualize citation relationships of publications. Hence citation graphs enable in-depth analysis about the impact of publications to research areas, such that citation graphs have great benefits for systematic reviews about a special field of research. In this contribution, we introduce a tool for automatically generating citation graphs from a set of paper documents, which runs stand-alone or integrated in a systematic reviews application. As systematic reviews often include many papers, we also propose several strategies to reduce the complexity of citation graphs and add additional information for in-depth analysis of the impact of single publications. In addition to citation graphs our tool also visualizes the publication selection process of systematic reviews. The generated graphs and developed strategies are evaluated using different instruments, including an user survey, in which they are rated positively.

TYPE OF PAPER AND KEYWORDS

Regular research paper: *citation graph, visualization, systematic review, publication, citation*

1 MOTIVATION

Even if we focus on special areas of research, there is an increasing number of publications and citations due to an exponential increase in publications [10] in general. Hence researchers and other interested readers have to deal with a flood of publications whenever they inform themselves in detail about the current state of research of a subject area. Systematic reviews [43] prepare the current state of research of a subject area in a way that help their readers to get a quick overview and understanding of this state of research. The term "systematic" is coined by the aim to actually find all the relevant and existing literature for a specific research question [20] in a systematic and standardized way.

A citation graph is a directed graph in which publications are the nodes and citations are the edges. By illustrating the citation relationships in citation graphs,

the spread of different methods, ideas or conceptions among the works like which works serve as basis for further work become visible [61], such that the impact of single publications and its authors can be determined. Citation graphs supports various other analysis in systematic reviews like identification of related research documents [39], ranking search results, trend analysis and social network analysis [7].

Hence citation graphs are a great kind of visualizations having benefits for readers of systematic reviews. However existing tools for systematic reviews provide only basic and very limited support for automatically generating visualizations for systematic reviews [1], such that the generation of citation graphs and even flow charts of the publication selection process presenting the flow of information during the different phases of the review are not covered at all. Hence we propose a tool called *ReViz* for automatically generating citation graphs

and flow charts of the publication selection process. ReViz can run stand-alone or smoothly integrated into a tool for systematic reviews.

Our contributions are:

- A tool for automatically generating flow charts of and citation graphs for systematic reviews.
- Our developed software for the creation of the visualizations (i.e., flow chart as well as citation graph) can be downloaded at <https://github.com/l-hartung/reviz> and is freely available to users worldwide by using docker containers.¹
- Different variants of citation graphs for simplifying the presentation by summarizing nodes and edges, and introducing factors like direct and indirect citations as well as coloring publications with common authors for further analysis.
- Extensive evaluations of the different variants of the citation graphs including
 - a quantitative analysis determining the effects of simplifying the presentation of citations graphs, and
 - a comprehensive user evaluation assessing the different variants of the citation graphs as well as the flow chart by external persons for the purpose of evaluating how well the results are understandable and appealing.

This contribution is an extended paper of [27] by detailing more about the basics, related work, requirements, architecture of the implementation and concepts, and by a more comprehensive evaluation including a detailed description of a user survey.

In the following Section 2, we introduce the basics of systematic reviews. In addition, previous work on the support of reviews as well as various visualizations and their selection for our work are presented. Subsequently, Section 3 details the concept: At first, requirements for the support of reviews and visualizations are discussed. Then we introduce methods and strategies for the flow chart and the citation graph with variants for simplification and adding information for the purpose of in-depth analysis. Afterwards in Section 4, we evaluate the created visualizations and Section 5 provides a summary and the future work.

2 BASICS

In this section, some basic principles are presented in order to shed more light on systematic reviews and

¹ In addition to the use of Parsifal and the export of data from it, it is also possible to generate a citation graph without creating a systematic review, using a Bibtext file and the referenced publications as PDF documents (local files or remotely accessible via urls) as input. However, the generation of a flow chart is only possible using Parsifal. Since this requires the use of a fork, the modified code of Parsifal is also available in the form of a docker container at <https://github.com/l-hartung/parsifal/>.

their preparation. In addition, some already existing tools supporting systematic reviews are considered in order to select one of these tools for the present work. Furthermore, different types of visualizations in reviews are introduced, including the flow chart of the publication selection process and the citation graph. Finally, the extraction of references from publications is discussed. This step is necessary for the implementation of the citation graph.

2.1 Systematic Reviews

Systematic reviews provide an overview of selected scientific papers on a research topic. Firstly, all available publications relevant to the research area are identified, evaluated and interpreted [32]. A central advantage of a systematic review over other scientific papers is its high informative value. While individual research papers are often based on the expectations of the scientists and results that do not correspond to the desired results can also be omitted from the publication, a systematic review is fundamentally more objective and very comprehensive. This means that all publications dealing with the question of the review must be included in the work, regardless of whether they support the desired hypothesis or not [32]. By summarizing a large number of existing research results on a topic, gaps, contradictions, relationships, or inconsistencies in the research can be identified, thus providing clues and directions for future research [32].

Originally, systematic reviews were mainly used in the field of medicine, after the British epidemiologist Archie Cochrane had a decisive influence on the methodology in the 1970s and 1980s [16] and in 1993 founded the *Cochrane Collaboration* for the preparation of medical reviews. Since then, reviews have been used to draw conclusions about medical interventions that a single study cannot provide. By considering all available studies in this field, much more precise conclusions about the intervention under investigation can be drawn. Thus, systematic reviews can provide evidence for the clinical relevance of different methods in the health care sector and help to develop guidelines for clinical decisions [43]. However, reviews have long since ceased to be a means in medicine; the same benefits are also drawn from them in the field of software engineering, so that more and more such reviews are being carried out there as well [1]. In principle, the application of reviews is possible in all research areas.

There are guidelines for the development of a review in order to create uniform and comparable results. In the medical field these are for example the *Cochrane Reviewer's Handbook* [53] and the *CRD Guidelines for those Carrying out or Commissioning Reviews*. [30] In

[32], these guides have been adapted for the research area of software engineering. In principle, however, the procedure of a review is mostly identical in every area, consisting of three different phases:

- In the planning phase, the need for a review in the respective specific research area should first be examined. The most important step in planning is then the *Review Protocol*. Here, all methods, strategies and criteria that will later be used in the search, selection and analysis of research work are planned and presented. Defining this *before* the beginning of the actual work is extremely important, for example to minimize *researcher bias*, i.e. distortions in the results, which arise unconsciously due to the expectations of the researchers [32].
- In the second phase, the actual conduct of the review, as much research as possible is initially collected using a defined search strategy. Subsequently, these must be examined in terms of their relevance to the research question. For this purpose, inclusion and exclusion criteria are used to categorize the work. In the later steps of the process, the quality of the remaining work is also assessed. For example, the extent to which attention was paid to minimizing various *bias* sources is taken into account. The remaining work after this step is followed by the extraction of the relevant data and their summary.
- In the third and final phase, the review is reported in a research report or in a journal or conference article.

2.1.1 PRISMA-Statement

As in other publications, the quality of a systematic review can vary greatly and depends to a large extent on the approach and thoroughness of the authors, as well as the scope and quality of the publications included. In order to ensure uniform standards, the PRISMA (*Preferred Reporting Items for Systematic reviews and Meta-Analyses*) statement was specified as a guideline for the report of systematic reviews [43]. Although the statement is designed for medical reviews, it can also serve as a basis for reviews from other areas. The PRISMA statement consists of a checklist of 27 points that should be included in the review report and a 4-phase flow chart. This is presented in Figure 1.

The *Preferred Reporting Items for Systematic reviews and Meta-Analyses for Protocols* (PRISMA-P) statement of 2015 provides guidance to facilitate the preparation of a protocol for systematic reviews [44]. According to Moher et al., far too few reviews make their protocol available to the reader, or create one at all. A detailed and conscientiously kept protocol is crucial for the traceability of the steps taken and thus also for the reproducibility and validity of the results of a review.

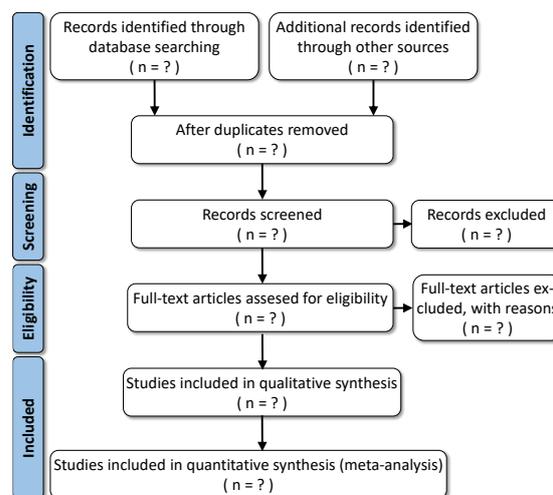


Figure 1: Flow chart to illustrate the flow of information during the different phases of a systematic review (adapted from [43])

The protocol thus plays a decisive role in the execution of a review. The fact that not all authors create one, however, indicates that support in the review creation process makes sense.

2.2 Related Work for Supporting Systematic Reviews

The extremely high effort required for the preparation of systematic reviews leads to an increased need for automatic support during this process. Many software tools are already available to help authors in this context. These range from basic word processing programs, *Reference Management Tools* and statistical programs to specially designed tools, which are intended to support the entire systematic review process - or large parts of it - [41]. *Reference Management Tools*, such as RefWorks and EndNote, are widely used by review authors [41]. Such tools are available in large numbers, but only take up a very small portion of the work in a review. The Cochrane Collaboration also offers several tools to support the management and analysis of systematic reviews, including Covidence [18], EPPI-Reviewer [55] and RevMan [54]. These tools are designed specifically for Cochrane medical reviews. However, Eppi-Reviewer and RevMan can also be used for other types of reviews, although in this case some features cannot be used, and provide good support in some areas of the systematic review process [41].

In [1] six other widely-used tools are compared and evaluated from different perspectives: Parsifal [47], StArt [28], SLuRp [11], SLR-Tool [25], SLRTOOL [6] and SESRA [45]. These six tools provide support

throughout the systematic review process and are not limited to a specific application area, although Al-Zubidy et al. refer to the software engineering area. The authors identify various barriers during the review process and requirements for supporting tools, and examine the six selected tools in relation to these. The tools cover different areas of the process to varying degrees. Although StArt, Parsifal, and SESRA provide the best overall results, there are also areas where they perform poorly. Overall, therefore, no tool could be found that could cover all the desired and required functionalities of the systematic review process.

Only Parsifal is free and open source software. The program is implemented as a web application in Python with Django, such that it can be extended for our purposes. In addition, Parsifal in [1] is among the three tools with the best overall results, so it provides a good basis for this proposed work.

There are other works that deal with the comparison of tools with respect to the requirements during the systematic review process, such as [40]. However, most of the tools either overlap with those already mentioned or support far less of the overall systematic review process. Accordingly, there are promising approaches to support systematic reviews; nevertheless, there are gaps in the existing tools, and optimal support cannot be provided. We provide a detailed discussion in Section 3.1 and Section 3.2 about the gaps and missing optimal support of systematic reviews in these existing tools.

2.3 Visualizations in Systematic Reviews

Visualizations in scientific works enable a better understanding of data sets, provide deeper insights or facilitate the analysis and presentation of large amounts of data [22]. Furthermore, in systematic reviews, different types of visualizations are used, for example to present information about the included publications in a clear way. Table 1 presents some types of visualizations frequently used in reviews and some sample reviews that include these visualizations.

Flow charts for the selection of the included publications during the entire review process are included in almost all of the reviews examined. It is part of the PRISMA statement and is displayed in all systematic reviews in the form specified there or in a very similar manner. Some of the systematic reviews display tables with the most important characteristics of the included publications to provide an overview of them. In most cases, different tables are also shown to illustrate the different properties and characteristics of the respective systematic review. Diagrams showing the distribution of the publication years of

the included publications are only used in the reviews examined in the area of software engineering. Other distribution diagrams like the publications sources or their geographical distribution, are also frequently found in this area. In medical reviews, on the other hand, forest plots are always presented to summarize and support their clues of the respective studies (e.g. relative risk or odds ratio). Funnel charts are also used here, for example, which can provide information on publication bias.

The most commonly used presentation is the flow chart of the publication selection process, which obviously forms a fundamental part of a review. Therefore it is implemented as one of the visualizations in this work to allow a quick and easy representation of the diagram in all systematic reviews.

2.3.1 Citation Graphs

Surprisingly none of the examined systematic reviews includes a citation graph, although citation analysis are the basis for tasks, which are very relevant for systematic reviews, like identification of related research documents [39], ranking search results, trend analysis and social network analysis [7]. Some tools for systematic reviews already cluster documents for the selection of publications, but they are currently clustered according to similarity of title, abstract and authors [46] due to missing citation data rather than according to citations promising good clustering results [7]. Overall this is an indication of the usefulness of citation analysis even in systematic review tools.

Definition 1 (Citation Graph): *A citation graph is a directed graph in which nodes represent publications and edges citations: An edge from node A to node B means that publication A cites B.*

Thus, the relationships between the publications are represented by the citations themselves [49]. An example of a citation graph is shown in Figure 2. With the analysis of citations and citation graphs, knowledge flows and the spread of ideas and perceptions as well as the relevance of information sources can be examined [61]. This can be of importance during the preparation of a systematic review. By illustrating the relationships of included publications in a graph, for example, the spread of different methods, ideas or conceptions among the works can be made visible. Furthermore, knowledge can be gained about the relevance of individual works and about which works serve as a basis for further work. The fact that a citation graph does not appear in any of the examined systematic reviews is therefore not necessarily an indication of its insignificance for systematic reviews, but could rather be an expression of the complexity and

Table 1: Various systematic literature reviews (SLR) and their applied visualizations. The first four reviews are from the medical area, the following four from the software engineering area and the last two from other areas.

SLR	Flow chart: selection of publi- cations	Table Study prop- erties	Other Tables	Distri- bution of publ.- years	Other distri- butions	For- est Plot	Fun- nel chart	Other
[26]	✓		✓			✓	✓	
[58]	✓		✓			✓		
[50]	✓					✓	✓	
[35]	✓	✓				✓		Risk of bias diagram
[60]	✓			✓	✓			classifications of relevant publications
[5]			✓	✓				
[8]	✓	✓	✓	✓				
[51]	✓		✓	✓	✓			
[21]	✓		✓					
[14]	✓	✓	✓					

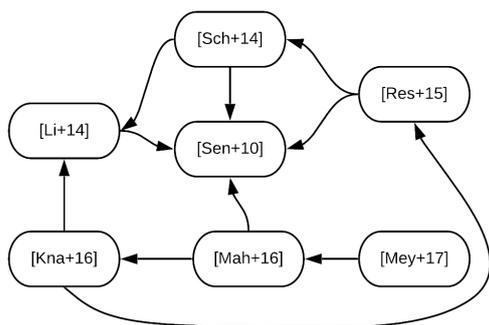


Figure 2: Example of a citation graph

efforts behind the creation of a citation graph. Therefore, a semi-automatic creation of citation graphs is a useful addition to the tool support of systematic reviews.

A citation graph can also allow such interpretation possibilities for the literature research of other works in which no systematic review is carried out. Thus, a citation graph for selected literature is generated for the present study, which is presented in Figure 3. There, especially in the middle part, many citations between publications about tools to support reviews are visible, so that it can be assumed that these works are based on similar ideas and that they build on each other.

2.3.2 Visualization Tools for Scholarly Datasets

Please see Table 2 for tools for visualizing scholarly datasets. Existing tools are not primarily designed for systematic reviews and offer many types of visualizations not necessarily including citation graphs suitable for systematic reviews. Hence our proposed tool ReViz offers some important unique features: Integrated in the tool Parsifal for conducting systematic reviews, the citation graphs of ReViz are automatically constructed (with possibility of manual correction) from a set of paper documents. ReViz further supports various simplification approaches for citation graphs.

2.4 Extraction of References

To generate a citation graph, the references from the publications of a systematic review must be examined for citations among each other. This requires an extraction of the references using a suitable tool. There are tools that are able to extract metadata from scientific publications. In most cases *linear chain condition random fields* (CRFs) [34] are applied [49]. These are used to create probability models for segmentation and labelling of sequential data. Examples of such tools are ParsCit [17] and Grobid [38]. Other tools like CiteSeerExtractor [59] perform data extraction in several parts, sometimes using other tools like ParsCit or Grobid.

In this paper we decide to use *GeneRation Of Bibliographic Data* (Grobid), which is a library for extracting metadata, such as bibliographic information,

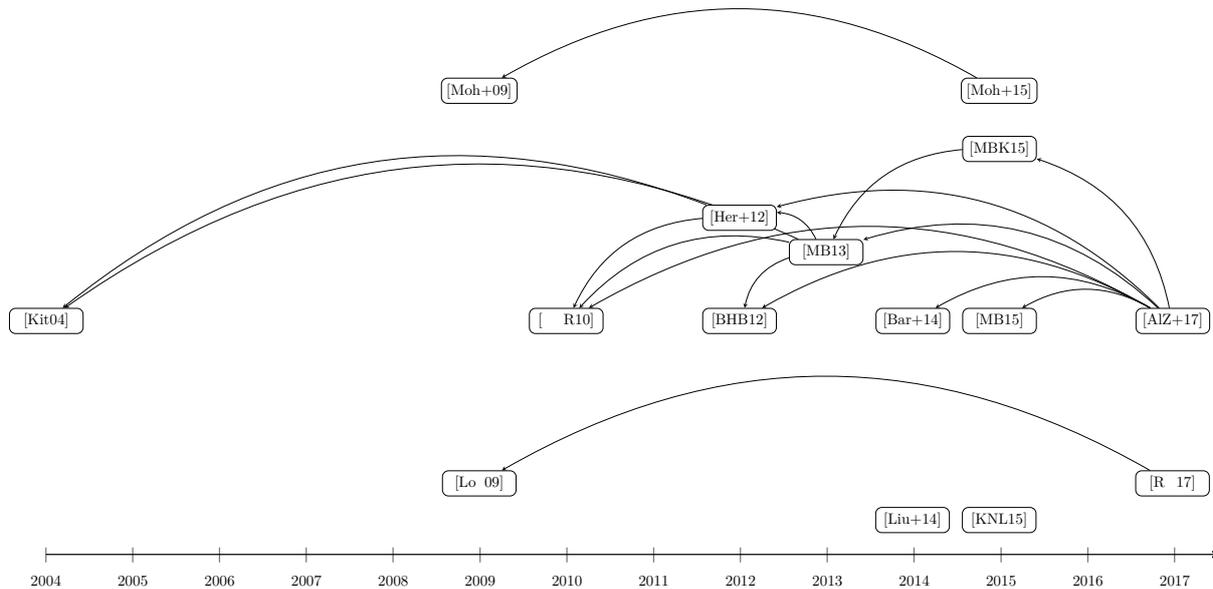


Figure 3: Citation graph for selected relevant literature of this contribution*

from scientific publications. It can be used to extract header data, references and full texts from articles in PDF format. The results can then be exported in whole or isolated (for example, individual references) to *TEI* files. *TEI* is an XML-based document format of the *Text Encoding Initiative*. Grobid offers both a Java API and a RESTful API. The program performs the classifications using CRFs and uses the Wapiti CRF library² for this purpose. In this paper, Grobid is used to filter out all existing references from the PDF files of the publications of a review. With this data it is then possible to find all citations of the publications among each other and use them for the citation graph.

3 FLOW CHART OF THE PUBLICATION SELECTION PROCESS, CITATION GRAPHS AND VARIANTS

We first discuss the requirements for the software support of systematic reviews with special regard to visualizations. Afterwards, we argue for flow charts and citation graphs as relevant types of visualizations suitable for further analysis of systematic reviews and we finally introduce several variants of citation graphs in this section.

² <https://wapiti.limsi.fr/>

* This citation graph is generated by our tool ReViz.

3.1 Requirements for the Software Technical Support of Systematic Reviews

The high effort that scientists spend for creating systematic reviews can be partly reduced by the use of supporting tools. Although there are steps of the systematic review in which only humans can carry out the work themselves, large areas of the creation process can be supported by software tools. In the following enumeration, some requirements for the creation of reviews, which can theoretically be supported by software, are listed (adapted and extended from [1]):

- Preparation of a review protocol with the search strategies, quality and selection criteria etc. to be used before starting the review
- Support in the search for publications; this step is extremely complex and can be facilitated, for example, by automatic search in different digital libraries, export of search results, standardization of search strings and removal of duplicates
- Support in the selection of relevant publications through textual analysis, search for similar publications and assistance in the management of inclusion and exclusion criteria
- Support for decisions on the quality of publications, for example by recording quality criteria during the preparation of the protocol and then querying them in phase 2 as a fillable form for each publication
- Automatic analysis and summary of the results from the publications
- Automatic creation of visualizations for the clear

Table 2: Various visualization tools for scholarly datasets

Visualization Tool	Visualizations*	Citation graph from set of paper documents	Support of systematic reviews	Comments
CitNetExplorer [57]	C	-	-	Clustering of very large citation networks.
VOSViewer [56]	B	-	-	Visualized bibliometric networks constructed based on citation, bibliographic coupling, co-citation, or co-authorship relations may include journals, researchers, or individual publications.
Sci²Tool [9]	B	-	-	Temporal, geospatial, topical, and network analysis and visualization of scholarly datasets at the micro (individual), meso (local), and macro (global) levels.
CiteSpace [13]	B	-	-	Structural and temporal analyses including collaboration networks, author co-citation networks, and document co-citation networks with support of hybrid node types such as terms, institutions, and countries, and hybrid link types such as co-citation, co-occurrence, and directed citing links.
CiteWiz [24]	C	-	-	Visualization of citation networks using causality visualization techniques, interactive timelines, and concept maps.
Dimensions [29]	C	-	-	Books, chapters, conference proceedings, grants, patents, clinical trials, policy documents and altmetric information are interlinked in a rich multipartite extension of a citation graph.
Convergent Clustering [12]	C	-	-	Citation data is used to identify and characterize putative colleges or communities of practice with special focus on small communities (representing new ideas).
Citation Trend Prediction [19]	C	-	-	Approach utilizing a graph neural networks-based architecture for predicting the top set of papers at the time of publication.
SciMAT [15]	B	-	-	Open source software tool developed to perform a science mapping analysis under a longitudinal framework supporting bibliometric networks like co-word, co-citation and bibliographic coupling, normalizations, clustering approaches, temporal and network analysis (including centrality and density) and complex measures like h-index, g-index, hg-index and q ² -index.
bibliometrix [3]	B	-	-	bibliometrix is an open source package for the programming environment R for bibliometric analysis and visualizations like country scientific collaborations, co-citation networks, keyword co-occurrences, co-word analysis and historical direct citation network.
Connected Papers [23]	S	-	-	Papers are arranged according to their similarity (based on co-citation and bibliographic coupling), i.e., even papers that do not directly cite each other can be strongly connected and very closely positioned.
Proposed tool ReViz	C	✓	✓	Tool especially designed for systematic reviews, and running stand-alone or integrated in Parsifal for visualizing citation graphs constructed from a set of paper documents. Support of various approaches for simplifying citation graphs.

* C: Focus on Citation Networks, B: General Bibliometric Network Visualizations, S: Clustering based on similarity

presentation of various information as well as for assistance in analyses and as a guide to the relationships of the included publications

- Management of the data, references, documents and publications included in the review
- Support for creating a review in a team, such that every participant has access to all data, can make changes and all changes are documented with the respective

author

Existing tools are previously discussed in Section 2.2. The six tools mentioned, which provide support for the entire review creation process, may already cover some of the requirements listed, but many of the requirements remain unmet:

To a large extent, the creation of a review protocol is supported by providing templates and requiring the

appropriate fields to be filled in before searching for publications. An automatic search for publications and further assistance in this area is only supported to a very limited extent. For the selection of relevant publications and assessment of quality, there are approaches in the tools, but there are still unexploited possibilities: For example, reports on statistical data as well as included and excluded publications can already be generated automatically for analysis. Furthermore, there are various visualizations which can be generated automatically in the individual tools. However, these are limited to a few diagrams that are relatively simple to generate, such as the distribution of publications according to their sources. Overall, the management of data for a review and its included publications is easier by using existing tools. Also, the work in a team can be coordinated very well.

For almost all of the listed requirements, support options are already available in the tools. However, most of them are still very superficial and limited in their possibilities. The first item, the review protocol, and the last two items, data management and team creation, are best covered. However, there is still a greater need for support for the other items.

3.2 Requirements for the Visualizations to be Generated

As already mentioned in Section 2.3, visualizations contribute, for example, to a better understanding of complex facts and to a better analysis of data. Visualizations in the context of systematic reviews can additionally support review authors in their work. There are some requirements for the visualizations to be generated in the context of a tool for systematic reviews. The following items deal with the requirements for promising visualizations for scientists and the possibilities in reviews that arise from these visualizations. Not all of the following requirements are covered by a single type of visualization. In the present work we hence select two types of visualizations, the flow chart and the citation graph, which we discuss here regarding the proposed requirements:

1. **Generation of a new view of the data, so that it becomes visible what would not be easily recognizable without visualizations:** This item is very general and can be achieved in principle by almost any meaningful visualization. For example, the flow chart provides insights into the selection of publications used and the number of articles discarded at the various stages of the review. Without this kind of visualization this can only be understood with considerably more effort. A citation graph also undoubtedly offers new perspectives on the

publications used and their relationships to each other.

2. **Representation of trends in the data, so that different developments, changes and different flows can be observed:** The presentation of trends in the data of a review is not easily possible in a statistical graph. A citation graph, however, provides a possibility for this representation. For example, different trends, i.e. the spread of knowledge in connected subgraphs of the citation graph, could be made visible. In this way, indications of different developments can be found in the relationships between the articles.
3. **Identify influences of a publication on other publications; for example, if the content, methods, ideas or developments of a publication originated from or are influenced by another publication:** Similar to the item above, influences of the publications among each other can certainly be made visible through the relationships in the citation graph. Even if a citation does not necessarily mean that the cited publication has a strong influence on the citing publication, it is in any case an indicator of this influence.
4. **Presentation of the quality of the included publications in relation to the review and its criteria, such that a visual comparison between the publications is possible:** Other types of visualizations are necessary to show the quality and fulfilled or unfulfilled criteria of the included publications. Neither in a flow chart nor in a citation graph is this well feasible. Only the number of citations is directly visible in the citation graph, which is an indicator for the overall quality of a publication.
5. **Illustration of the selection process of the publications, the sources used, criteria violated and publications discarded:** A flow chart presents the selection process of the publications as an overview. It can show the libraries used and the number of articles found in the libraries, as well as the number of selected publications due to duplicates, violated criteria or insufficient importance for the systematic review.
6. **Presentation and grouping of important characteristics of the included publications:** In principle, neither the flow chart nor the citation graph is suitable for depicting many characteristics of the publications. However, to a certain extent it is possible to present additional information in a citation graph. This can be achieved by suitable stylistic means, such as a timeline for the graph to show the publication years of the individual articles.

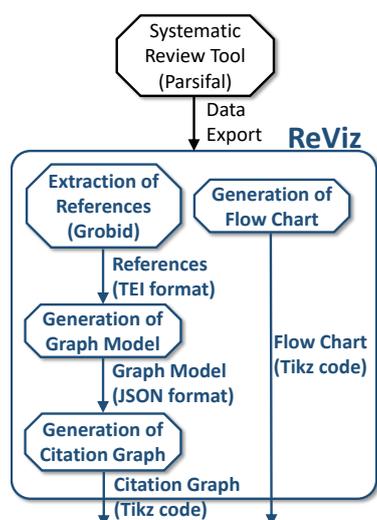


Figure 4: Pipeline for creating the citation graph and the flow chart

3.3 Architecture

The creation of the selected visualizations is divided into several steps, the pipeline of which is presented in Figure 4. A software tool to support systematic reviews³ provides the data about the included publications in the systematic review, which are needed for the visualizations. Subsequently, this data is used to create the flow chart. On the other hand, the references of the publications are also extracted for the citation graph to obtain the necessary information about the edges in the graph. Thus, we develop a data model for the graph, which is used in the following step to create the citation graph. The individual components of the pipeline should represent separate steps that can be executed independently of each other so that the pipeline does not have to be run through completely each time for individual results.

3.4 Flow Chart of the Publication Selection Process

Our proposed visualization generator (integrated into Parsifal) generates a flow chart for the publication selection process (see Figure 5), which is based on the structure of the template contained in the PRISMA statement [43]. In our study of several reviews on different topics, we observed differences in the presentation of the flow charts, but the content is always based on the PRISMA statement. In these flow charts, nodes present publications found in various digital libraries or other sources. Furthermore, nodes for the

³ Actually, we use Parsifal in our implementation.

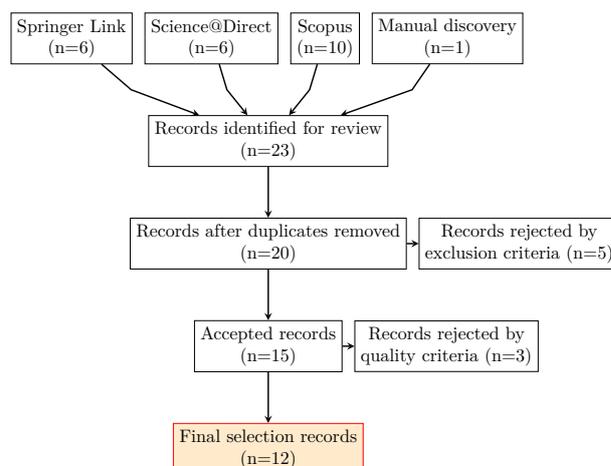


Figure 5: Example flow chart of the publication selection process of systematic reviews. This flow chart is generated by our tool ReViz.

number of publications after duplicate elimination, as well as after removal of publications by inclusion and exclusion criteria and quality criteria should be included. Thus, each step in the publication selection process of the review is shown in a summarized form.

3.5 Graph Model for Citation Graphs

The data model of the citation graph should contain all basic elements needed to build the graph. These are primarily all occurring publication years, information about the included publications as well as information about the citations of the publications among themselves. To calculate the edges, all references of the publications must be compared with the publications actually used in the systematic review in order to find mutual citations. For this purpose, DOIs of the articles as well as titles and authors are a good means for comparison. Figure 6 presents a flow chart demonstrating the procedure of the comparison.

Whenever DOIs are available for both the reference and the publication, the DOIs are compared: If they are identical, there is no need to look any further at the two articles under investigation, since the DOIs are unique and therefore a citation, i.e. an edge in the citation graph to create, can already be identified at this point. If there is only one or no DOI or if the DOIs available are not identical, the titles are considered in the next step. If the titles match, the authors are also checked (in order to achieve a reliable result). Otherwise, it is tested whether the titles are at least approximately the same so that relatively similar titles are not directly discarded. In this case the user has to decide whether reference and publication match. Even if the authors do not match in

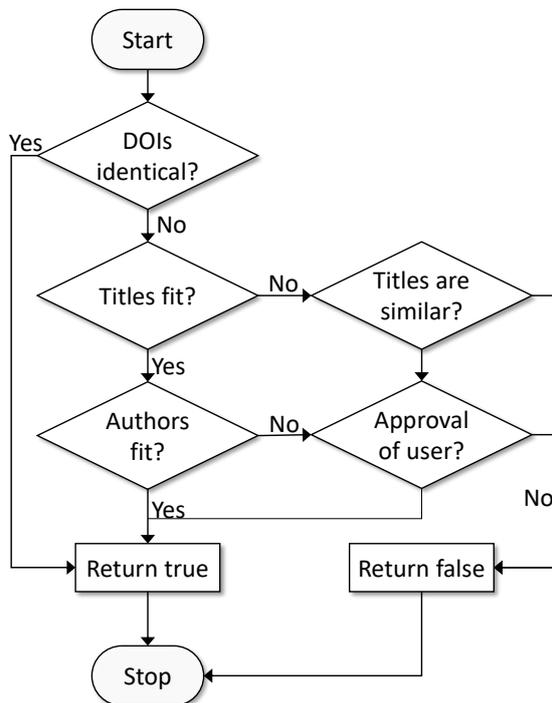


Figure 6: Program flow chart to compare the references with the included publications

the first step, the user decides on the result. Thus, cases in which the program cannot make a reliable statement are passed on to the human being. On the basis of this decision, an edge is then identified accordingly or the processing is continued with the next articles.

3.6 Structure of the Citation Graphs

Citation graphs in the context of systematic reviews should provide an insight into the relationships between the publications included in the review. The aim is to create a meaningful graph for the reader that provides as much information as possible. In order to achieve this goal, a further component is included here in our graphs: the arrangement of the publications should be based on their respective year of publication. This means that all publications which appeared in the same year are arranged next to each other, making additional information visible. For illustration purposes, a timeline is printed below the graph itself, so that it is possible to see exactly which publication appeared in which year. In addition, it can be quickly determined in which period of time the included research work is conducted and from which years more or less work originates.

The basis for the citation graph is therefore primarily the timeline. It covers the period of all publication years of the included works. Based on the timeline, the

works are then drawn as nodes at the respective position on the timeline and the works linked by quotations are connected to each other with edges. Furthermore, different components within the graph should be visibly separated from each other. From a graph-theoretical point of view, it is possible that the citation graph is not only represented by a single directed graph, but consists of several independent subgraphs. In the context of this work, however, it is useful to consider the citation graph always as a single unit and thus as a graph with several *components*. Here, a component is defined as a subset of nodes and edges in which each node has at least one incoming or outgoing edge to another node of this subset. In addition, each node that has no edges is also its own component. To make the independence of the components visible, they should be drawn one below the other.

3.7 Node Summaries

Although various methods are used here to obtain an illustrative result, the large number of nodes and edges in a citation graph can still lead to very confusing results. An example is shown in Figure 7. In order to reduce this complexity, further methods to simplify the graph are necessary.

One way to reduce overlaps of many edges and the resulting confusion is to reduce the number of nodes. A smaller number of nodes results in fewer edges and a less dense overall graph. If many publications are included in a review, the only way to reduce the number of nodes is to merge several individual nodes. In this respect, a summary of the original graph is generated. The difficulty in creating small graph summaries is the minimization of the resulting errors [36], so that no information is missing from the original graph and no wrong information is added.

There are various approaches to summarizing graphs, but they are not necessarily applicable to such relatively small graphs like our citation graphs. A merging of several nodes to a *supernode* with *superedges* is a well realizable possibility. Such a merging is demonstrated in Figure 8. Several nodes with the same or very similar incoming and outgoing edges are merged into one large node. In the case of the citation graph, this one large node contains then all publications, which influence and are influenced by (about) the same publications. Publications of the same super node seem to often address exactly the same research question and/or their contributors are from the same community (having the same research background and having the same knowledge in terms of state-of-the-art). Hence, investigating the publications of the super nodes can be the starting point for these kinds of analysis in

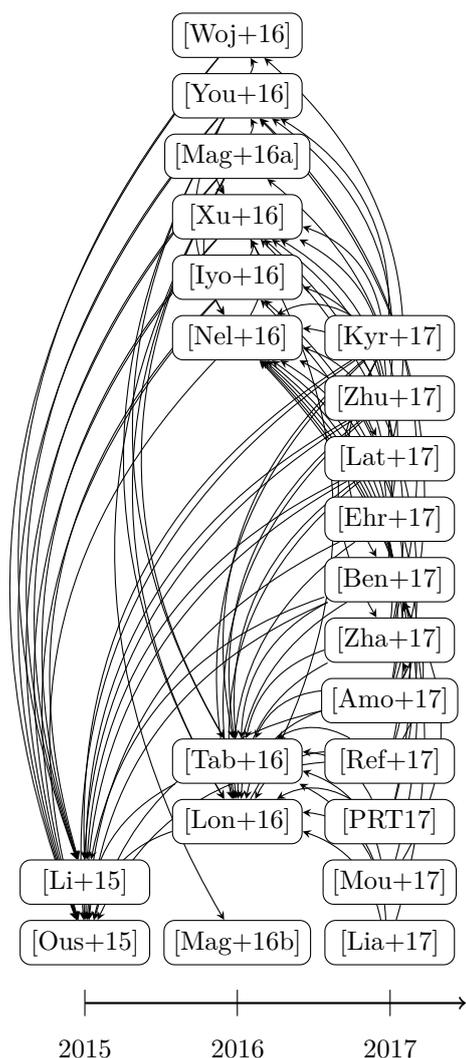


Figure 7: Example of a citation graph (for the systematic review presented in [4])*

systematic reviews as well. In order to minimize errors and to reconstruct the original graph exactly, the merge consists not only of the graph itself, but also of a list of corrections of the edges.

This method can quickly become very inefficient for large graphs, since numerous comparisons of the nodes with each other must be processed to find those with similar edges. In [31] such an algorithm is presented, where first similar nodes are searched for to avoid unnecessary comparisons between all nodes. Then a summary of the graph is iteratively generated by merging original nodes or already existing *supernodes*.

In case of the citation graphs, only merges of nodes in the same level, i.e. publications with the same publication year, are reasonable. Because of this,

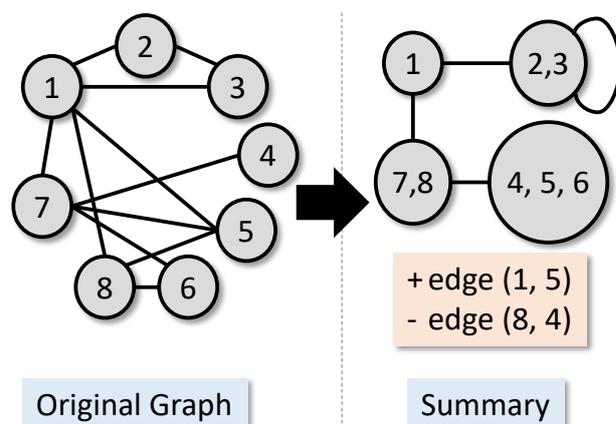


Figure 8: A graph and its summary after merging nodes after merging several nodes with respective corrections of the edges (adapted from [37])

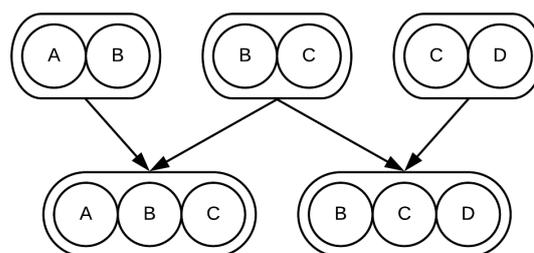


Figure 9: A connected component consisting of three candidate pairs and the resulting candidate for the fusion of three nodes

supernodes with a large number of merges are extremely unlikely, so it makes more sense to focus on good merges of two or three nodes each. For this purpose, all possible candidates for the merging of two nodes are found first. Using a weight based on common and different edges for each pair of candidates, the best possible candidates are then selected. Afterwards, possible combinations of the candidates for the fusion of three nodes must be examined. This is done by the calculation of connection components from the candidate pairs. Figure 9 illustrates this approach.

The upper row shows a possible connected component of three candidate pairs. It should be noted here that not all combinations of candidates in a component can necessarily form a triple combination; the candidate with articles A and B cannot be combined with the last candidate with articles C and D. The possible triple combinations of the candidates of the context component are shown in the bottom row of the figure. A weight is

also determined for such triple candidates, so that the best candidates can ultimately be determined.

We present the above described example in Figure 10 as a summary after merging nodes: A total of 5 nodes and 37 edges are removed.

3.8 Summarizing Transitive Edges

If the publications in the graph are scattered over a longer period of time, there are inevitably fewer ways to merge nodes, since there are far fewer nodes in a plane. We present an example in Figure 11.

In this case, the clarity is impaired by many very long edges. To avoid this, more edges would have

to be reduced, regardless of the number of nodes. Transitivity is suitable for this purpose. The term transitivity originates from set theory and is defined as follows:

Let R be a relation on a set A , i.e. $R \subset A \times A$. R is called *transitive*, if $(a, b) \in R$ and $(b, c) \in R$ is followed by $(a, c) \in R$, where $a, b, c \in A$ [42].

Here the set A corresponds to the nodes, while R contains the directed edges. We present in Figure 12 an example of a transitive edge.

Node A has both an edge to C , as well as to node B , which in turn is also connected to C by an edge. If one regards the edges as an information flow, redundant information is transmitted; two different paths lead from A to C . By omitting the direct edge from A to C the knowledge about the connection of the two nodes is not lost, it remains indirectly through the connection via B . Similarly, in the citation graph, these transitive edges can be left out to reduce the amount of edges without withholding essential information. This does not only apply to connections via three nodes, but can run through any number of edges. In Figure 13, in addition to the edge from A to C , the edge from A to D can also be omitted.

Figure 14 presents the result of the citation graph after leaving out the transitive edges. Compared to the original graph, 8 edges less are drawn in this figure.

3.9 Other Citation Graph Variants

After the summary of transitive edges, some information is lost in the graph. For example, the number of incoming edges is no longer a clear indication of the number of citations of this node. In order to keep this information, we propose to increase the width of the remaining incoming edges of a node for omitted transitive edges (see Figure 15). Thus, depending on the size of the incoming edges, the actual number of citations can be better inferred.

As a further variant and in order to have a metric for the influence of single publications to other research contributions, we propose to display exact numbers in the nodes of the citation graphs for (direct) citations as well as for indirect ones. Indirect citations represent a path of direct citations (see Figure 16), i.e., A indirectly cites B if A cites C and C indirectly cites B , where A , B and C are publications. This allows a direct comparison between the citations of the individual nodes, despite the omitted edges. In order not to enlarge the nodes too much by the two additional numbers in the label, they must be displayed relatively small. In order to enable a quick comparison of the quotations of the individual nodes at first glance, these numbers are additionally highlighted in color. By means of a

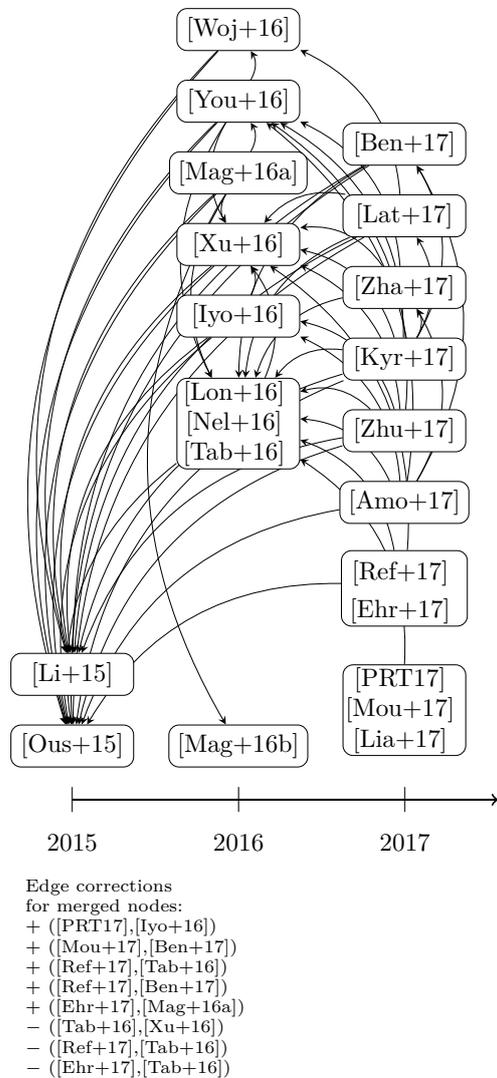


Figure 10: Example citation graph with merging of several nodes and correction of the edges (for the systematic review presented in [4])*

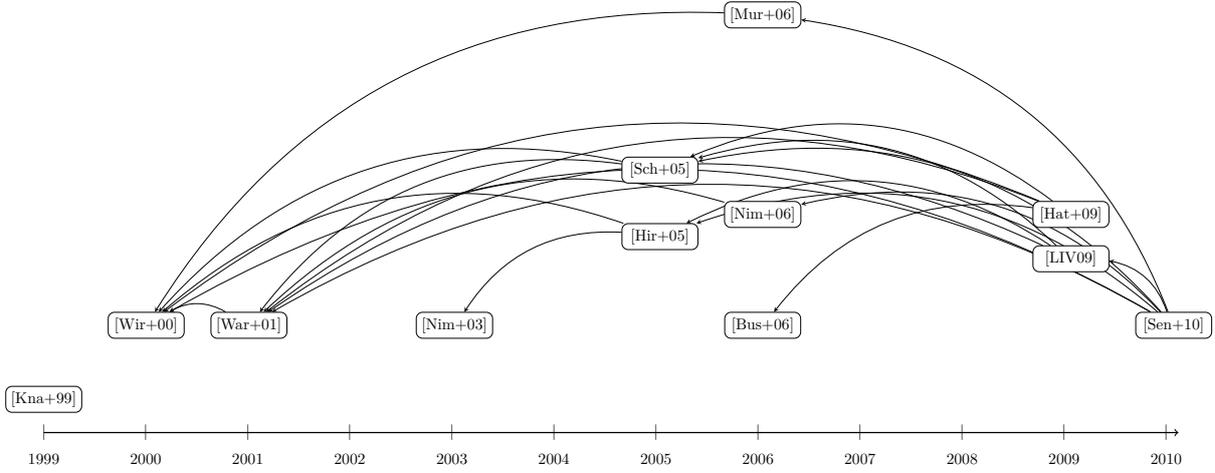


Figure 11: Example of a citation graph covering a longer time period (for the systematic review presented in [33])*

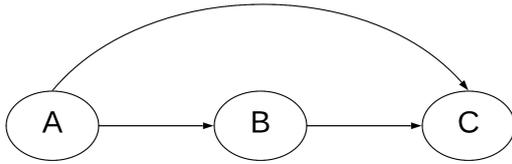


Figure 12: Example of a transitive edge

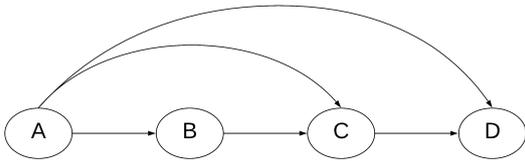


Figure 13: Example of transitive edges over multiple edges

color scale, nodes with many citations can be quickly distinguished from those with fewer citations.

Additionally, it should be possible to identify publications with many common authors, for example to recognize follow-up publications and related approaches more easily. We hence propose to draw these publications in the same color (see Figure 17). However, the number of authors in different publications can vary considerably and the number of common authors is likely to carry more weight if the total number of authors is smaller. In addition, the first author of a publication also has significantly more influence on it. For this reason, the following calculation is used to compare the

authors of two publications:

Publication 1 has the authors $A = \{a_1, a_2, \dots, a_n\}$,
 Publication 2 has the authors $B = \{b_1, b_2, \dots, b_m\}$

$$match(A, B) = \sum_{i=1}^n \sum_{j=1}^m \begin{cases} 5 : & a_i = b_j, i = j = 1 \\ 3 : & (a_i = b_1, i \neq 1) \\ & \vee (a_1 = b_j, j \neq 1) \\ 1 : & a_i = b_j, i, j \neq 1 \\ 0 : & otherwise \end{cases}$$

$$score(A, B) = \frac{match(A, B)}{|A \cup B|}$$

The value $score(A, B)$ takes into account the total number of authors as well as first authors among the joint authors and weights them higher. A value of 0 indicates no joint authors. If it is normalized to the interval $[0, 1]$ larger than a threshold value to be defined by the user, the two publications can be colored in the same color. Because of the normalization, the choice of the threshold value depends on the distribution of the $score$ for all pairs of articles. For many citation graphs it is reasonable to color publications with one (or more) joint author(s).

Furthermore, since not only two publications can have common authors, additional candidates with common authors must be found among these candidates, so that it is also possible to mark three or more publications with the same color.

4 EVALUATION

The evaluation of the visualizations introduced in this work mainly focuses on different aspects of the citation

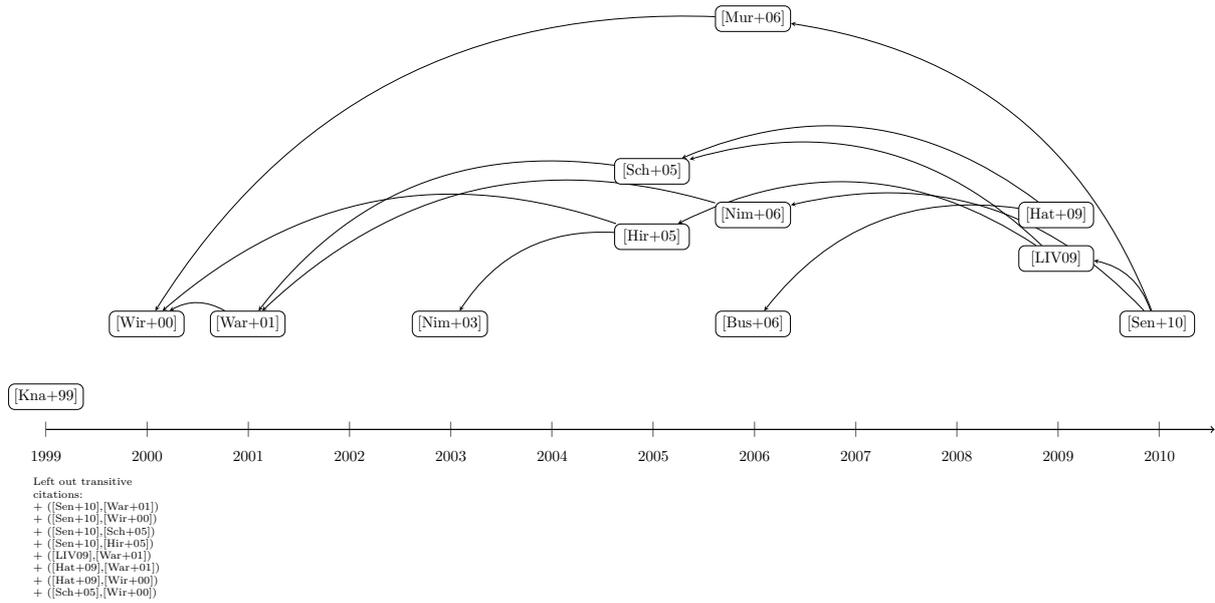


Figure 14: Example citation graph after removing transitive edges*

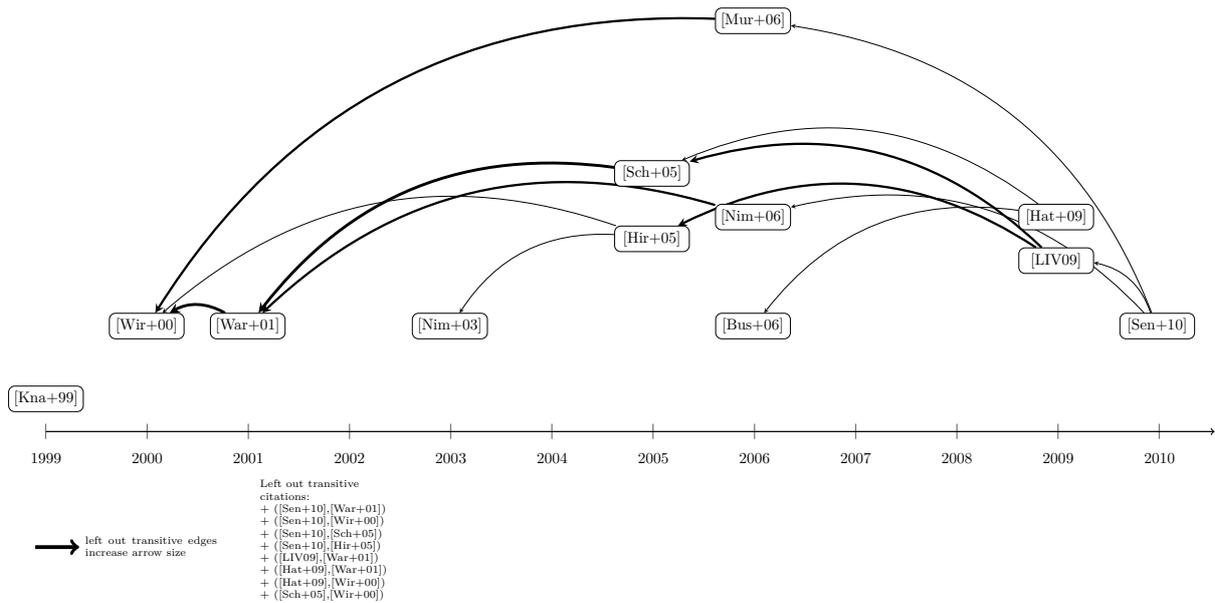


Figure 15: Example citation graph after removing transitive edges with increasing width for more removed transitive edges*

graph. The developed flow chart, which is largely based on the 4-phase flow chart of the PRISMA statement, offers little scope for variation and therefore takes up only a small part of the evaluation.

For the citation graph, on the other hand, several decisions are made regarding the layout and the used elements. However, the evaluation of the results is complex and can only partly be calculated

mathematically. The manual of TikZ comments on this issue:

There is no “perfect” way of drawing a graph, rather, depending on the circumstances there are several different ways of drawing the same graph and often it will just depend on the aesthetic sense of the reader which layout he or she would prefer. ([52, S.411])

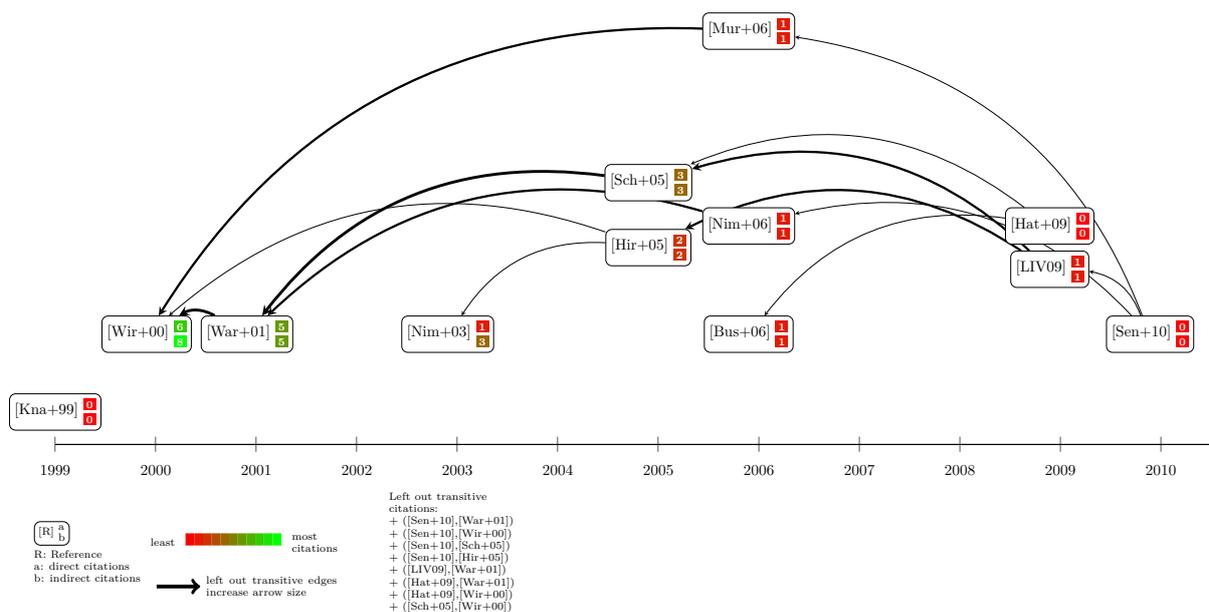


Figure 16: Example citation graph after removing transitive edges with direct and indirect citations numbers*

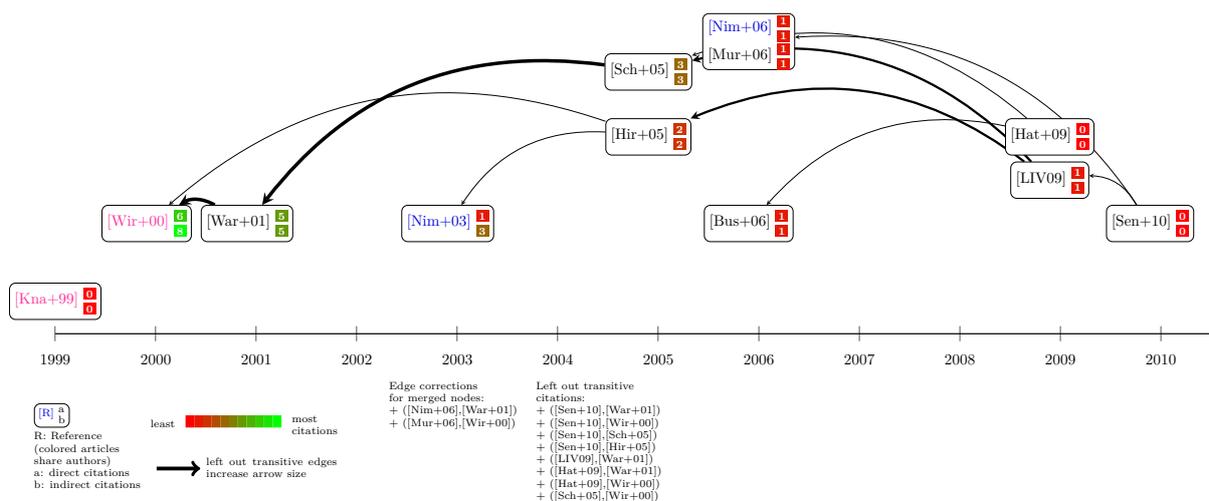


Figure 17: Example citation graph after removing transitive edges with summarized nodes and with publications colored for joint authors*

Since the evaluation of the visualizations is partly subject to the subjective perceptions of the reader, part of the evaluation consists of an user survey. Thus, different aspects of the created variants of the citation graph as well as the flow chart can be evaluated. In addition to the survey, some calculations regarding the number of nodes and edges in the citation graph are carried out for evaluation. Furthermore, the performance of reference recognition by Grobid and the subsequent assignment of these references to included papers is manually validated.

4.1 Identifying Edges

The basis for the citation graph is formed by the citations of the publications among themselves. Only if these can be correctly identified by Grobid and our tool, the graph displays all information correctly. Checking the correctness of the identified edges and finding missing edges, however, is a relatively high effort, since all references of the included publications must be examined manually. For this reason, two examples (see figures 18 and 19 for the citation graphs for the systematic reviews in [2] and [33] respectively) with 12

Table 3: Results of the evaluation of two examples to identify edges

	Example 1 (Figure 18 and [2])	Example 2 (Figure 19 and [33])
True positive	11	21
False negative	0	3
False positive	0	0

publications each, one with slightly fewer and one with slightly more edges, are inspected for the evaluation in this context. The results are given in the table 3.

In the first example, all 11 existing citations are correctly identified and no errors are made by our tool. Two of the citations had to be confirmed by the user due to missing or incorrect authors in the data. In the second example a large fragment of the edges is found by our tool and also here each of the found edges is correct. None of the 21 identified edges required the user's approval, so that our proposed algorithm already allows to assume correct edges with sufficient certainty. However, three existing citations are not identified by our tool. It should be noted here that all three are the same cited publication and are not listed in the TEI files issued by Groid. Therefore, all references listed in the TEI files are processed correctly by our tool. In this respect, it can be assumed that Groid has difficulties in extracting individual references in seldom cases so that they cannot be identified.

4.2 Reduction of the Complexity of Citation Graphs

The number of nodes and edges can vary greatly in the introduced variants of citation graphs. Fewer nodes and edges reduce the complexity in the graph. We present in Table 4 the results for different calculations regarding nodes and edges for two graphs in different variants.

The upper part of the table contains the number of nodes, edges, and edges per node. Overall, there is a very strong reduction of the evaluated values. In the first graph, only about half of the edges of the original graph are drawn using both the node summary and the summary of transitive edges, while in the second example, there is even a reduction of 73% in total edges and 67% in edges per node.

In the lower part of the table, three additional parameters are considered that make it difficult to track edges: the number of nodes hiding one or more edges, the number of edges hidden by nodes, and the number of edge overlaps. Since these parameters are sometimes difficult to measure visually and could only be examined

manually, some of the values are approximate values. Here, too, the results with node summary and summary of transitive edges are to a large extent highly reduced. In the case of edge overlaps, even 90% and 95% lower values can be achieved in both graphs when using both functionalities (i.e., node summary and removing transitive edges).

Long edges, which run over a longer span on the timeline and thus across several layers, add to the confusion, as they are more likely to cross more other edges and nodes, making it more difficult to quickly capture the course of all edges. Therefore, in Figure 20, we present the lengths of edges occurring in the two graphs considered earlier, so that a comparison of the edge lengths in the graph variants is possible. In addition to the general reduction in the number of edges, which is already shown in Table 4, it is also apparent that many of the longer edges are eliminated by removing transitive edges.

Overall, very high reductions for the evaluated values occur in the calculations presented here, whenever node summaries and summaries of transitive edges are performed in the graphs. Thus, a lower complexity of the graphs can be concluded. In order to determine whether better final results for the graphs can be achieved as a consequence, the results of the user survey follow in the next section.

4.3 User Survey

In an user survey created for the evaluation of the visualizations, the different implemented variants of the citation graphs as well as the flow chart are assessed by external persons. This allows to determine how well the results are understandable and appealing to outsiders and whether the desired goals for the visualizations have been achieved. The graphs are primarily examined with regard to the points "clarity" (*Are all information quickly and easily grasped at a glance?*), "comprehensibility" (*Are all necessary information available to understand the overall picture?*) and "layout" (*Is the result visually appealing?*).

By means of the survey, the following hypotheses regarding the created variants of the citation graphs and flow chart are to be tested:

1. Any implemented functionality for the citation graph is useful.
2. Comprehensibility, clarity and layout depend on the number of nodes and edges in the citation graph.
3. Every implemented functionality for the citation graph improves clarity and layout.
4. The comprehensibility of the citation graph does not change by using the different implemented functionalities.

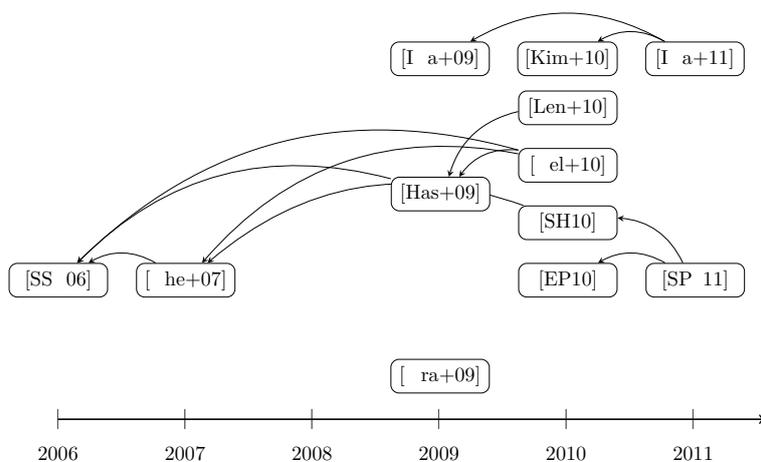


Figure 18: Example 1 for the evaluation of the identification of the edges (citation graph of the systematic review in [2])*

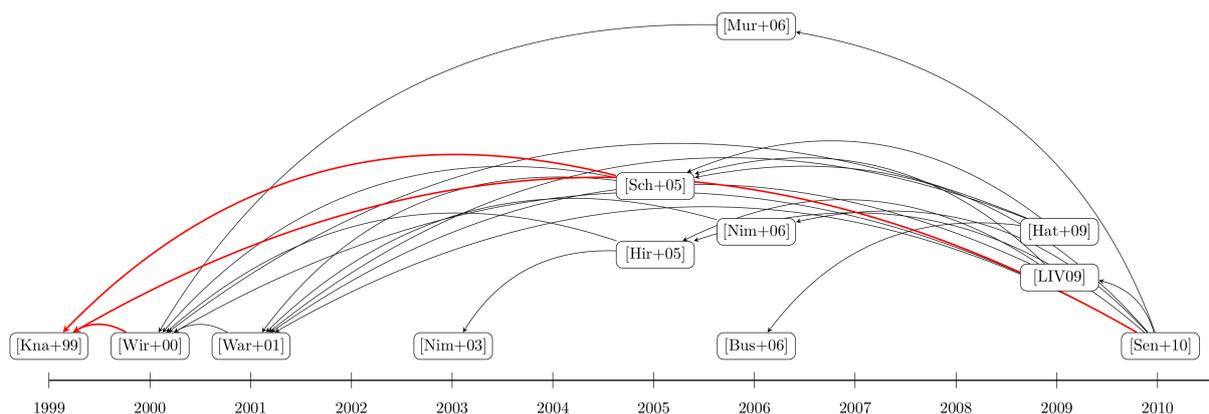


Figure 19: Example 2 for the evaluation of the identification of the edges, unidentified edges (false negative) are shown in red (citation graph of the systematic review in [33]). This citation graph has been generated by our tool ReViz, where the unidentified edges have been manually added.

5. Analyses of the importance of a publication and the influence on other publications are easier to perform due to the changed edges thickness with transitive edges left out and the colored highlighting of the number of citations.
6. The flow chart is clear, comprehensible and visually appealing.
7. It makes sense to present the generated flowchart in the review in addition to the textual description.

The first part of the survey records the employment and experiences level of the participants (see Figure 21). A total of 22 people participated in the survey, mostly research assistants and students. The majority of

them are well versed in literature research and are also familiar with systematic reviews. Thus a large part of the participants form the target group for the visualizations introduced in this work.

The individual questions on the visualizations are always to be answered with a rating of 1 to 5, with 1 being the lowest and 5 the best rating. In the following paragraphs, the hypotheses listed above and their respective sections of the survey are considered individually.

Hypothesis 1: The first hypothesis is examined in the survey by assessing each of the individual functions

Table 4: Results for the number of nodes, edges and edges per node, as well as the number of nodes hiding edges, edges hidden by nodes and edge overlaps for two graphs (Graph 1 contains the publications of [33], Graph 2 contains the publications of [4]). In the first line the original – normal – graph is taken as starting point. This is followed by the graph with a summary of similar nodes, where an edge deviation of two has been allowed for nodes to be combined, and the graph with a summary of transitive edges. In the last line, both functionalities are combined.

		Nodes	Edges	Edges per Node
Graph 1 [33]	Normal Graph	12	21	3,5
	1. Node Summary	11 (-8%)	18 (-14%)	3,27 (-7%)
	2. Transitivity	12 (0%)	13 (-38%)	2,17 (-38%)
	1. & 2.	11 (-8%)	10 (-52%)	1,81 (-48%)
Graph 2 [4]	Normal Graph	22	92	8,36
	1. Node Summary	18 (-18%)	60 (-35%)	6,67 (-20%)
	2. Transitivity	22 (0%)	40 (-57%)	3,64 (-56%)
	1. & 2.	18 (-18%)	25 (-73%)	2,78 (-67%)
		Hiding Nodes	Hidden Edges	Overlaps
Graph 1 [33]	Normal Graph	4	8	~42
	1. Node Summary	4 (0%)	10 (+25%)	33 (-21%)
	2. Transitivity	3 (-25%)	2 (-75%)	7 (-83%)
	1. & 2.	3 (-25%)	2 (-75%)	4 (-90%)
Graph 2 [4]	Normal Graph	18	~77	~220
	1. Node Summary	14 (-22%)	~47 (-39%)	~90 (-59%)
	2. Transitivity	18 (0%)	~32 (-58%)	~44 (-80%)
	1. & 2.	13 (-28%)	~19 (-75%)	10 (-95%)

of the implemented citation graphs according to its usefulness. Figure 22 illustrates the results graphically. Each functionality is assigned with a color and the numbers of the respective answers from 1 to 5 are shown in the bar chart. Even if the answers differ partly, it is clear that the majority of the results for each functionality are in the range of 4 and 5 and can therefore be classified as meaningful. Only very few answers rated individual functions in the range of 1 and 2, i.e. as not useful.

Hypothesis 2: In the survey, three citation graphs, in which the number of nodes and edges and the time period of the included publications vary greatly, are shown. For each of these graphs, clarity, comprehensibility and layout should be evaluated. Figure 23 presents the 95% confidence intervals of responses for each of these criteria of the three graphs. The mean value is additionally marked. Graph 1 of the survey presents the highest values for all three criteria, with a relatively low variance. For the other two graphs, slightly more variance can be observed in the recorded responses. Nevertheless, the confidence intervals in the individual criteria do not overlap with those of the first graph,

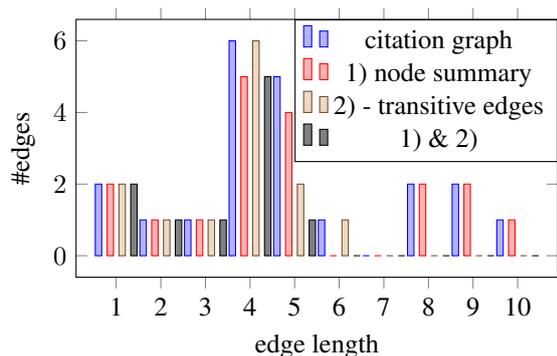
but are located much further down in the evaluation. Also between graphs 2 and 3 there is only a very small overlap between the confidence intervals for the layout. Otherwise, the answers to the three criteria for the different graphs differ very strongly. It can therefore be shown that the second hypothesis is correct for the confidence level selected.

Hypothesis 3: To consider the third hypothesis, different sections of the survey must be examined. One section of the survey assesses a graph using the summarization of nodes in relation to the original graph. The answers provide a 95% confidence interval of [4.00, 4.45] for clarity and [3.36, 4.10] for layout. In this context, a value of 3 would mean an equal evaluation of both graphs. Since in both cases the 3 is not included in the interval, the hypothesis for this functionality is proven. Table 5 presents the confidence intervals for each of the five functionalities.

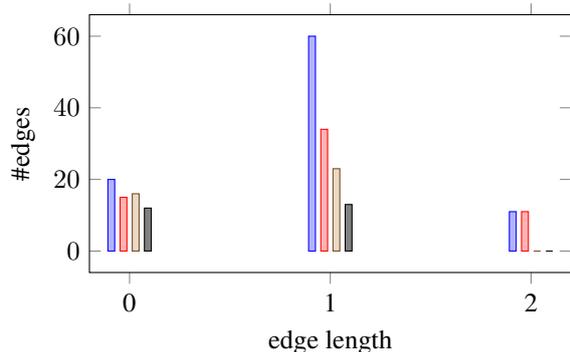
Hypothesis 4: Similar to hypothesis 3, the fourth hypothesis is tested for the comprehensibility of the citation graph for the different functionalities and a confidence interval is calculated for each (see Table 6).

Table 5: Results of the 95% confidence intervals for improving the clarity and layout of the graphs using the different functionalities. The intervals were limited upwards to 5.

Functionality	Clarity	Layout
Node Summary	[4.00, 4.45]	[3.36, 4.10]
Removing Transitive Edges	[4.50, 5.00]	[4.00, 4.64]
Edge thickness of transitive edges	[3.49, 4.24]	[3.27, 4.10]
Colored Number of Citations	[3.21, 4.25]	[3.38, 4.35]



(a) Graph 1



(b) Graph 2

Figure 20: Comparison of the edge lengths of a) Graph 1 and b) Graph 2 without any reduction, using node summary, removing transitive edges and both functionalities. The edge length on the x-axis indicates how many layers (i.e., years) an edge passes over.

In this case, the confidence intervals are somewhat closer to the three in relation to the upper ones, but only the interval for the node summary includes them just about. Overall, the hypothesis is thus refuted and a change in the comprehensibility of the graphs is observed. However, this change is limited to an improvement throughout, so that the participants of the survey are of the opinion that the citation graphs are easier to understand when using the different functionalities.

Hypothesis 5: A 95% confidence interval of [4.08, 4.56] is calculated for the fifth hypothesis, whereby here the ratings 4 and the 5 represent a high evaluation and thus an improvement in the analysis possibilities. Thus the hypothesis is proven.

Hypothesis 6: Hypothesis 6 is proven by the survey with a confidence interval of [3.68, 4.32].

Hypothesis 7: In addition to the example of a flow chart, we present the textual description of the systematic survey process so that the participants can directly compare the graphical presentation with the text form. The answers regarding the clarity, comprehensibility and layout of the diagram resulted in the 95% confidence intervals [4.31, 5.00], [4.31, 5.00] and [3.88, 4.75], whereby the upper limit was set at 5. The seventh hypothesis is hence clearly proven.

Hypothesis 8: The meaningfulness of the flowchart is also confirmed with a confidence interval of [3.71, 4.66].

4.3.1 Discussion of the Results

Almost all of our proposed hypotheses are proven by the survey. With 22 participants, the sample is relatively small and the answers of the participants varied from one another (see for example Figure 22). Nevertheless, a clear tendency in the answers can already be determined here. Since the confidence intervals include the variance of the answers and these are often nevertheless relatively far above 3 and thus in the upper range of the evaluation, the sample is large enough for the representativeness of the statements.

The variance in the responses with respect to the variants of the citation graph illustrates the subjective feelings of the users. In this respect, it is positive that the variants of the citation graph can be selected by the user himself. Thus, it is possible to apply the desired functionalities and to vary them according to the corresponding citation graph and the preferences of the user.

Table 6: Results of the 95% confidence intervals for the comprehensibility of the graphs using the different functionalities

Functionality	Comprehensibility
Node Summary	[2.99, 3.74]
Removing Transitive Edges	[3.78, 4.18]
Edge thickness of transitive edges	[3.73, 4.45]
Colored Number of Citations	[3.42, 4.22]

In general, the clarity of the two citation graphs, which contain more edges and/or nodes, is rated as relatively poor. As confirmed by the evaluation, the clarity is improved by the different functionalities. However, even after the improvement, a “good” result cannot necessarily be assumed. In order to be able to combine many nodes and thus save many nodes and edges, a high number of nodes within one year is necessary. In this case, however, the graph is likely to be very confusing, so that the result is likely to be complex even after the summary is performed. If, on the other hand, fewer nodes are within a year, but are scattered over a longer period of time, there are fewer possibilities to summarize, so that the clarity of the result will also not change much. However, if you use the summary of transitive edges, you often save a lot of edges, which certainly has a positive effect on the clarity of the graphs. Nevertheless, much of the information is obscured and the presentation of the omitted edges in the legend enables to trace the citations of a publication without gaps, but it is very time-consuming. Nevertheless, both functions are considered useful and are in any case capable of creating a new, clearer, presentation method for many citation graphs.

The use of the flow chart created by our tool is also an useful addition to the creation of a systematic review. Both hypotheses regarding the diagram are confirmed by the survey.

5 SUMMARY AND CONCLUSIONS

We introduce a tool for the creation of visualizations for systematic reviews. In particular, we integrated the generation of a flow chart for the publication selection process of the systematic review and different variants of citation graphs (with and without merging nodes, removing transitive edges, adding numbers for direct and indirect citations and coloring publications with common authors) for the analysis of the citations of the publications among each other. The basis for these visualizations is the data from the tool *Parsifal*, which supports the creation process of a systematic review. For the citation graph the tool *Grobid* is additionally

used to extract the references from the PDF files of the publications. We verify good results in an extensive evaluation by determining and comparing the number of hidden edges and hiding nodes in the citation graph variants and by a user survey for assessing subjective opinions of users.

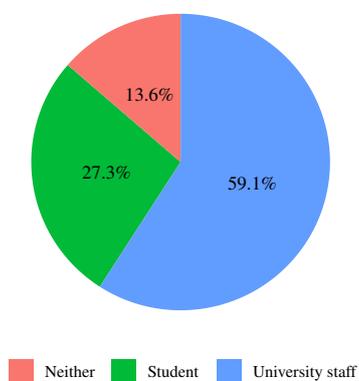
The Sugiyama algorithm, a popular method for arranging the nodes in a directed graph, is used to construct the citation graph. Our tool, however, generates also a timeline in the graph, which also ensures that the nodes are ordered according to the respective years of publication. In this respect, an innovative approach for the generation of citation graphs is developed.

The number of publications included in reviews is extremely variable and often 100 or more publications are included. In such cases, the generated citation graph is not useful and, due to the complexity, will in all likelihood not provide any advantages for the analysis of the publications. Nevertheless, there are enough reviews of topics specialised to a narrow topic or recent trends with a smaller number of publications for which such an application can be beneficial. In addition, the citations of the publications vary among themselves depending on the topic, so that a relatively large number of publications with relatively few citations among themselves can still result in a clear and promising citation graph.

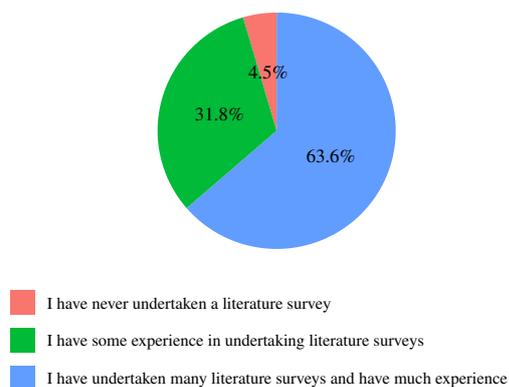
Furthermore, we discuss variants of citation graphs summarizing nodes and removing transitive edges for simplifying large citation graphs (without losing any or not much information), and marking publications of common authors with the same colors for further analysis. In this way also large citation graphs for systematic reviews dealing with many references can be displayed and analyzed in a reasonable way.

Independent of the number of publications used in the review or their citations, however, the flow chart of the publication selection process is generated quickly and easily with the developed tool, in addition to working out the review in *Parsifal*. In any case, the use of *Parsifal* is a useful way of simplifying the process of creating the review, since according to Al-Zubidy et al. it is one of the three best tools available to support the overall

(a) Participants' background



(b) Participants' level of experiences in literature surveys



(c) Participants' level of experiences in systematic literature surveys

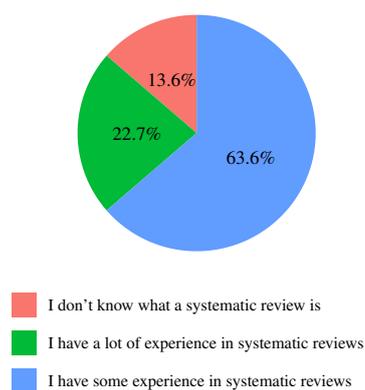


Figure 21: Level of experiences and employment of survey participants

process of a systematic review [1], although there is still

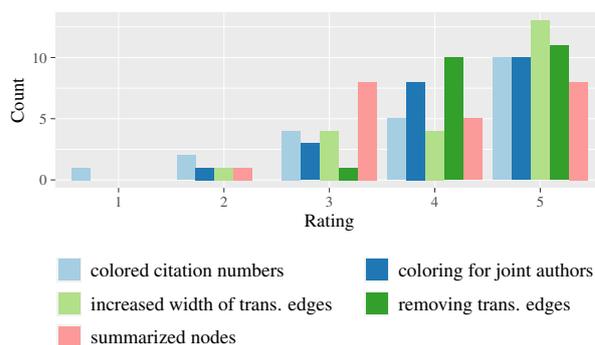


Figure 22: Results on the usefulness of the individual functionalities of the citation graph, whereby the rating 1 was determined as *not useful at all* and the rating 5 as *very useful*

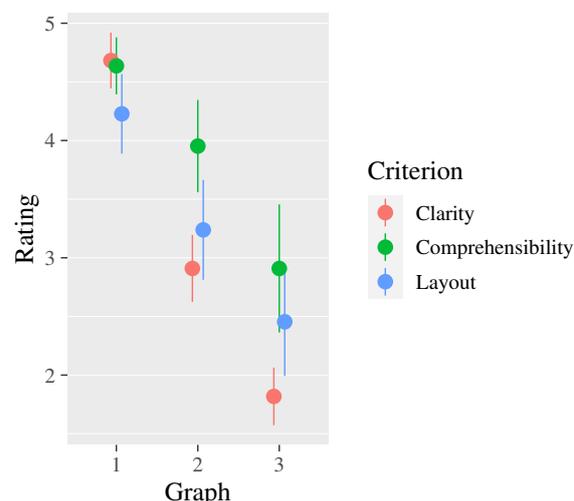


Figure 23: Display of the 95% confidence intervals for the answers on clarity, comprehensibility and layout of three different graphs with varying number of nodes and edges

some untapped potential for further support especially for generating visualizations.

One possible way to extend the developed citation graph in the future work is to add interactive elements, such as highlighting relevant information using “Mouse Over”. In this way, it would be possible to trace individual quotations in more complex graphs by highlighting only the incoming or outgoing edges of a node or by hiding all other edges when selecting a node. Although such interactive elements cannot be used everywhere, there are enough possibilities when displaying citation graphs on (review accompanying) web pages, for own use for analyses or similar.

ReViz is published as open source and is virtualized in docker containers for easy deployment at any computer. For the use of the program, the implementation of a graphical user interface is a useful extension, since it often allows for easier execution when using the tool locally. Manual changes to the output would also be easier to perform in this case. Furthermore, the presentation as a web service would allow the tool to be used without installation. This would also make it easier to integrate the visualizations into new or existing tools to support systematic reviews.

One of the most important steps in creating the citation graph is the identification of citations between the participating publications. Although the extraction of citations by Grobid has yielded very good results, the use of ontology-based approaches (of the Semantic Web) is an interesting way to supplement this. A promising project in this context is *OpenCitations* [48] for collecting citations of scientific publications, which in November 2019 comprises almost 14 million references and makes them available via corresponding Semantic Web technologies.

With the two types of discussed visualizations, the range of visualization possibilities for supporting reviews is not yet exhausted. It would also be conceivable, for example, to create visualizations that depict the quality of the included publications according to the defined quality criteria. In a *Heatmap*, which compares the publications and the criteria in two dimensions, a visual comparison of the quality could be made possible by using colors. The generation of visualizations for further meta-analyses in the reviews is another support option. This is already partly possible in the tools for supporting medical Cochrane reviews, so that for example *Forest Plots* can be generated automatically. However, meta-analyses in reviews are also conceivable in other areas, so that this possibility could be supplemented here for tools such as Parsifal.

REFERENCES

- [1] A. Al-Zubidy, J. C. Carver, D. P. Hale, and E. E. Hassler, "Vision for slr tooling infrastructure: Prioritizing value-added requirements," *Information and Software Technology*, vol. 91, pp. 72 – 81, 2017.
- [2] N. Alsufyani, C. Flores-Mir, and P. Major, "Three-dimensional segmentation of the upper airway using cone beam ct: a systematic review," *Dentomaxillofacial Radiology*, vol. 41, no. 4, pp. 276–284, 2012.
- [3] M. Aria and C. Cuccurullo, "bibliometrix: An r-tool for comprehensive science mapping analysis," *Journal of Informetrics*, vol. 11, no. 4, pp. 959 – 975, 2017. [Online]. Available: <https://doi.org/10.1016/j.joi.2017.08.007>
- [4] H. Babačić, A. Mehta, O. Merkel, and B. Schoser, "Crispr-cas gene-editing as plausible treatment of neuromuscular and nucleotide-repeat-expansion diseases: A systematic review," *PLoS one*, vol. 14, no. 2, 2019.
- [5] H. F. Badawi, F. Laamarti, and A. El Saddik, "Iso/ieee 11073 personal health device (x73-phd) standards compliant systems: A systematic literature review," *IEEE Access*, vol. 7, pp. 3062–3073, 2019.
- [6] B. Barn, F. Raimondi, L. Athiappan, and T. Clark, "Slrtool: a tool to support collaborative systematic literature reviews," in *ICEIS 2014 - 16th International Conference on Enterprise Information Systems*, 2014.
- [7] L. Bolelli, S. Ertekin, and C. L. Giles, "Clustering scientific literature using sparse citation graph analysis," in *European Conference on Principles of Data Mining and Knowledge Discovery (PKDD)*, 2006.
- [8] R. Bonidia, L. A. L. Rodrigues, A. P. Avila-Santos, D. Sanches, and J. Brancher, "Computational intelligence in sports: A systematic literature review," *Adv. Human-Computer Interaction*, vol. 2018, pp. 3 426 178:1–3 426 178:13, 2018.
- [9] K. Börner and D. E. Polley, *Visual insights: A practical guide to making sense of data*. MIT Press, 2014.
- [10] L. Bornmann and R. Mutz, "Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references," *Journal of the Association for Information Science and Technology*, vol. 66, no. 11, pp. 2215–2222, 2015.
- [11] D. Bowes, T. Hall, and S. Beecham, "Slurp: A tool to help large complex systematic literature reviews deliver valid and rigorous results," in *Proceedings of the 2Nd International Workshop on Evidential Assessment of Software Technologies*, ser. EAST '12. ACM, 2012, pp. 33–36.
- [12] S. Chandrasekharan, M. Zaka, S. Gallo, T. Warnow, and G. Chacko, "Finding scientific communities in citation graphs: Convergent clustering," *arXiv preprint arXiv:2007.14452*, 2020.
- [13] C. Chen, "Citespace II: detecting and visualizing emerging trends and transient patterns in scientific literature," *J. Assoc. Inf. Sci. Technol.*, vol. 57,

- no. 3, pp. 359–377, 2006. [Online]. Available: <https://doi.org/10.1002/asi.20317>
- [14] F. Chen, A. M. Lui, and S. M. Martinelli, “A systematic review of the effectiveness of flipped classrooms in medical education,” *Medical education*, vol. 51, no. 6, pp. 585–597, 2017.
- [15] M. Cobo, A. López-Herrera, E. Herrera-Viedma, and F. Herrera, “Scimat: A new science mapping analysis software tool,” *Journal of the American Society for Information Science and Technology*, vol. 63, no. 8, pp. 1609–1630, 2012. [Online]. Available: <https://doi.org/10.1002/asi.22688>
- [16] A. L. Cochrane, *Effectiveness and Efficiency. Random Reflections on Health Services*. Nuffield Provincial Hospitals Trust, 1972.
- [17] I. G. Councill, C. L. Giles, and M.-Y. Kan, “Parscit: an open-source crf reference string parsing package.” in *LREC*, vol. 8. International Language Resources and Evaluation, 2008, pp. 661–667.
- [18] Covidence Team, “Covidence.” [Online]. Available: <https://www.covidence.org>
- [19] D. Cummings and M. Nassar, “Structured citation trend prediction using graph neural networks,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain*, 2020.
- [20] H. de Vet, A. Verhagen, I. Logghe, and R. Ostelo, “Literature research: Aims and design of systematic reviews,” *Australian Journal of Physiotherapy*, vol. 51, no. 2, pp. 125–128, 2005. [Online]. Available: [https://doi.org/10.1016/s0004-9514\(05\)70041-6](https://doi.org/10.1016/s0004-9514(05)70041-6)
- [21] H. De Vries, V. Bekkers, and L. Tummers, “Innovation in the public sector: A systematic review and future research agenda,” *Public administration*, vol. 94, no. 1, pp. 146–166, 2016.
- [22] D. DiBiase, A. MacEachren, J. Krygier, and C. Reeves, “Animation and the role of map design in scientific visualization,” *Cartography and Geographic Information Science - CARTOGR GEOGR INF SCI*, vol. 19, pp. 201–214, 10 1992.
- [23] A. T. Eitan, E. Smolyansky, and I. K. Harpaz, “Connected papers.” [Online]. Available: <https://www.connectedpapers.com/>
- [24] N. Elmqvist and P. Tsigas, “Citewiz: A tool for the visualization of scientific citation networks,” *Information Visualization*, vol. 6, no. 3, pp. 215–232, 2007.
- [25] A. Fernández-Sáez, M. Genero, and F. Romero, “Slr-tool - a tool for performing systematic literature reviews.” in *ICSOFT 2010 - Proceedings of the Fifth International Conference on Software and Data Technologies*, vol. 2, 01 2010, pp. 157–166.
- [26] N. B. Finnerup, N. Attal, S. Haroutounian, E. McNicol, R. Baron, R. H. Dworkin, I. Gilron, M. Haanpää, P. Hansson, T. S. Jensen, P. R. Kamerman, K. Lund, A. Moore, S. N. Raja, A. S. C. Rice, M. Rowbotham, E. Sena, P. Siddall, B. H. Smith, and M. Wallace, “Pharmacotherapy for neuropathic pain in adults: a systematic review and meta-analysis,” *The Lancet Neurology*, vol. 14, no. 2, pp. 162 – 173, 2015.
- [27] S. Groppe and L. Hartung, “Reviz: A tool for automatically generating citation graphs and variants,” in *22nd International Conference on Asia-Pacific Digital Libraries (ICADL), Online*, 2020.
- [28] E. Hernandez, A. Zamboni, S. Fabbri, and A. D. Thommazo, “Using gqm and tam to evaluate start-a tool that supports systematic review,” *CLEI Electronic Journal*, vol. 15, no. 1, pp. 3–3, 2012.
- [29] D. W. Hook, S. J. Porter, and C. Herzog, “Dimensions: Building context for search and evaluation,” *Frontiers in Research Metrics and Analytics*, vol. 3, p. 23, 2018. [Online]. Available: <https://www.frontiersin.org/article/10.3389/frma.2018.00023>
- [30] K. S. Khan, G. Ter Riet, J. Glanville, A. J. Sowden, J. Kleijnen *et al.*, *Undertaking systematic reviews of research on effectiveness: CRD’s guidance for carrying out or commissioning reviews*. NHS Centre for Reviews and Dissemination, 2001, no. CRD Report 4 (2nd edition).
- [31] K. U. Khan, W. Nawaz, and Y.-K. Lee, “Set-based approximate approach for lossless graph summarization,” *Computing*, vol. 97, no. 12, pp. 1185–1207, Dec 2015.
- [32] B. Kitchenham, “Procedures for performing systematic reviews,” *Keele, UK, Keele University*, vol. 33, no. 2004, pp. 1–26, 2004.
- [33] P. L. Kubben, K. J. ter Meulen, O. E. Schijns, M. P. ter Laak-Poort, J. J. van Overbeeke, and H. van Santbrink, “Intraoperative mri-guided resection of glioblastoma multiforme: a systematic review,” *The lancet oncology*, vol. 12, no. 11, pp. 1062–1070, 2011.
- [34] J. Lafferty, A. McCallum, and F. C. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data.”

- Proceedings of the Eighteenth International Conference on Machine Learning, 2001, pp. 282–289.
- [35] J. Liu, C. Hua, J. Pan, B. Han, and X. Tang, “Piezosurgery vs conventional rotary instrument in the third molar surgery: A systematic review and meta-analysis of randomized controlled trials,” *Journal of Dental Sciences*, vol. 13, no. 4, pp. 342–349, 2018.
- [36] X. Liu, Y. Tian, Q. He, W.-C. Lee, and J. McPherson, “Distributed graph summarization,” in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, ser. CIKM ’14. New York, NY, USA: ACM, 2014, pp. 799–808.
- [37] Y. Liu, T. Safavi, A. Dighe, and D. Koutra, “Graph summarization methods and applications: A survey,” *ACM Comput. Surv.*, vol. 51, no. 3, pp. 62:1–62:34, Jun. 2018.
- [38] P. Lopez, “Grobid: Combining automatic bibliographic data recognition and term extraction for scholarship publications,” in *International conference on theory and practice of digital libraries*. Springer Berlin Heidelberg, 09 2009, pp. 473–474.
- [39] W. Lu, J. Janssen, E. Milios, N. Japkowicz, and Y. Zhang, “Node similarity in the citation graph,” *Knowledge and Information Systems*, vol. 11, no. 1, 2007. [Online]. Available: <https://doi.org/10.1007/s10115-006-0023-9>
- [40] C. Marshall and P. Brereton, “Tools to support systematic literature reviews in software engineering: A mapping study,” in *International Symposium on Empirical Software Engineering and Measurement*, 10 2013, pp. 296–299.
- [41] C. Marshall, P. Brereton, and B. Kitchenham, “Tools to support systematic reviews in software engineering: A cross-domain survey using semi-structured interviews,” in *Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering*, 04 2015, pp. 26:1–26:6.
- [42] F. Modler and M. Kreh, *Tutorium Analysis 1 und Lineare Algebra I*. Springer Spektrum, 2011.
- [43] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and the PRISMA Group, “Preferred reporting items for systematic reviews and meta-analyses: The prisma statement,” *Annals of Internal Medicine*, vol. 151, no. 4, pp. 264–269, 08 2009.
- [44] D. Moher, L. Shamseer, M. Clarke, D. Ghersi, A. Liberati, M. Petticrew, P. Shekelle, and L. A. P. Stewart, “Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015 statement,” *Systematic Reviews*, vol. 4, no. 1, p. 1, 01 2015.
- [45] J. S. Molléri and F. B. V. Benitti, “Sesra: A web-based automated tool to support the systematic literature review process,” in *Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering*, ser. EASE ’15. ACM, 2015, pp. 24:1–24:6.
- [46] M. Ouzzani, H. Hammady, Z. Fedorowicz, and A. Elmagarmid, “Rayyan—a web and mobile app for systematic reviews,” *Systematic Reviews*, vol. 5, no. 1, p. 210, 2016. [Online]. Available: <https://doi.org/10.1186/s13643-016-0384-4>
- [47] Parsifal Ltd., “Parsifal.” [Online]. Available: <https://parsif.al/>
- [48] S. Peroni, A. Dutton, T. Gray, and D. Shotton, “Setting our bibliographic references free: towards open citation data,” *Journal of Documentation*, vol. 71, no. 2, pp. 253–277, 2015.
- [49] F. Rosenthal and S. Groppe, “Purposeful searching for citations of scholarly publications,” *Open Journal of Information Systems (OJIS)*, vol. 4, no. 1, pp. 27–48, 2017. [Online]. Available: <http://nbn-resolving.de/urn:nbn:de:101:1-201711266882>
- [50] M. Shalaby, S. Emile, H. Elfeki, A. Sakr, S. D. Wexner, and P. Sileri, “Systematic review of endoluminal vacuum-assisted therapy as salvage treatment for rectal anastomotic leakage,” *BJS open*, vol. 3, no. 2, pp. 153–160, 2018.
- [51] S. Soheilirad, K. Govindan, A. Mardani, E. K. Zavadskas, M. Nilashi, and N. Zakuan, “Application of data envelopment analysis models in supply chain management: a systematic review and meta-analysis,” *Annals of Operations Research*, vol. 271, no. 2, pp. 915–969, 12 2018.
- [52] T. Tantau, “Graph drawing in tikz,” in *Graph Drawing*, W. Didimo and M. Patrignani, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 517–528.
- [53] The Cochrane Collaboration, “Cochrane handbook for systematic reviews of interventions,” 2011. [Online]. Available: <http://handbook.cochrane.org>
- [54] The Cochrane Collaboration, *Review Manager (RevMan) [Computer Program]*, version 5.3 ed., Copenhagen: The Nordic Cochrane Centre, 2014.
- [55] J. Thomas, S. Graziosi, and J. Brunton, “Eppi-reviewer.” [Online]. Available: <http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2914>

- [56] N. J. van Eck and L. Waltman, "Software survey: Vosviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523–538, 2010. [Online]. Available: <https://doi.org/10.1007/s11192-009-0146-3>
- [57] 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. [Online]. Available: <https://doi.org/10.1016/j.joi.2014.07.006>
- [58] P. F. Whiting, R. F. Wolff, S. Deshpande, M. Di Nisio, S. Duffy, A. V. Hernandez, J. C. Keurentjes, S. Lang, K. Misso, S. Ryder, S. Schmidkofer, M. Westwood, and J. Kleijnen, "Cannabinoids for medical use: A systematic review and meta-analysis," *JAMA*, vol. 313, no. 24, pp. 2456–2473, 06 2015.
- [59] K. Williams, L. Li, M. Khabsa, J. Wu, P. C. Shih, and C. L. Giles, "A web service for scholarly big data information extraction," in *2014 IEEE International Conference on Web Services*, June 2014, pp. 105–112.
- [60] J. Yli-Huumo, D. Ko, S. Choi, S. Park, and K. Smolander, "Where is current research on blockchain technology? — a systematic review," *PloS one*, vol. 11, no. 10, p. e0163477, 2016.
- [61] D. Zhao and A. Strotmann, "Analysis and visualization of citation networks," *Synthesis Lectures on Information Concepts, Retrieval, and Services*, vol. 7, no. 1, pp. 1–207, 2015.

AUTHOR BIOGRAPHIES



Sven Groppe earned his diploma degree in Computer Science in 2002 and his Doctor degree in 2005 from the University of Paderborn. He earned his habilitation degree in 2011 from the University of Lübeck. His publication record contains about 100 publications including the book

"Data Management and Query Processing in Semantic Web Databases" published by Springer. He was a member of the DAWG W3C Working Group, which developed SPARQL. He was the project leader of the DFG project LUPOSDATE, an open-source Semantic Web database, and of two research projects in the area of FPGA acceleration of relational and Semantic Web databases. He is also leading a DFG project on GPU and APU acceleration of main-memory database indexes, and a DFG project about Semantic Internet of Things. He is also the chair of the Semantic Big Data workshop series, which is affiliated with the ACM SIGMOD conference (so far 2016 to 2020), and of the Very Large Internet of Things workshop in conjunction with the VLDB conference (so far 2017 to 2020). He is general chair of the upcoming International Semantic Intelligence Conference (ISIC) in 2021. His research interests include artificial intelligence, databases, Semantic Web, query and rule processing and optimization, Cloud Computing, acceleration via GPUs and FPGAs, peer-to-peer (P2P) networks, Internet of Things, data visualization and visual query languages. Detailed information is available at <https://www.ifis.uni-luebeck.de/~groppe/>.



Lina Hartung earned her Bachelor of Science in Medical Informatics at the University of Lübeck in 2017 and completed her studies with the Master's degree in 2019. A major focus of her studies was on achieving semantic interoperability of (health care) systems, which she examined in her Bachelor's thesis. Her Master's thesis, titled "Software Assistance

for Generating Visualisations in Systematic Literature Reviews", serves as the basis of this paper.