Visualizing Research Impact through Citation Data

YONG WANG, the Hong Kong University of Science and Technology
CONGLEI SHI, Airbnb Inc.
LIANGYUE LI and HANGHANG TONG, Arizona State University
HUAMIN QU, the Hong Kong University of Science and Technology

Research impact plays a critical role in evaluating the research quality and influence of a scholar, a journal, or a conference. Many researchers have attempted to quantify research impact by introducing different types of metrics based on citation data, such as *h*-index, citation count, and impact factor. These metrics are widely used in the academic community. However, quantitative metrics are highly aggregated in most cases and sometimes biased, which probably results in the loss of impact details that are important for comprehensively understanding research impact. For example, which research area does a researcher have great research impact on? How does the research impact change over time? How do the collaborators take effect on the research impact of an individual? Simple quantitative metrics can hardly help answer such kind of questions, since more detailed exploration of the citation data is needed. Previous work on visualizing citation data usually only shows limited aspects of research impact and may suffer from other problems including visual clutter and scalability issues. To fill this gap, we propose an interactive visualization tool, *ImpactVis*, for better exploration of research impact through citation data. Case studies and in-depth expert interviews are conducted to demonstrate the effectiveness of *ImpactVis*.

CCS Concepts: • **Human-centered computing → Information visualization**; *Visualization design and evaluation methods*;

Additional Key Words and Phrases: Research impact, publication and citation, visualization

ACM Reference format:
Yong Wang, Conglei Shi, Liangyue Li, Hanghang Tong, and Huamin Qu. 2018. Visualizing Research Impact through Citation Data. *ACM Trans. Interact. Intell. Syst.* 8, 1, Article 5 (March 2018), 24 pages.
https://doi.org/10.1145/3132744

5

1 INTRODUCTION

Research impact, in a figurative sense, represents the influence that a scholar, a journal, or a conference poses on others. In the academic community, research papers of a scholar can be cited by other researchers, and this can be naturally used to indicate that these papers have an influence

The reviewing of this article was managed by special issue associate editors Yu-Ru Lin and Nan Cao.
This work is supported in part by the National 973 Program of China (2014CB340304) and RGC GRF 618313. Hanghang Tong is supported by NSF IIS-1651203, DTRA HDTRA1-16-0017, ARO W911NF-16-1-0168, NIH R01LM011986, and a Baidu gift.
Authors' addresses: Y. Wang and H. Qu, Department of Computer Science and Technology, the Hong Kong University of Science and Technology, Hong Kong, China; emails: {ywangct, huamin}@cse.ust.hk; C. Shi, Airbnb Inc., San Francisco, USA; email: shiconglei@gmail.com; L. Li and H. Tong, School of Computing, Arizona State University, Tempe, USA; emails: {liangyue, hanghang.tong}@asu.edu.

https://doi.org/10.1145/3132744

on the researchers who cited them. Finding a proper way to evaluate such kind of impact is very important for many application scenarios. For example, research funding bodies pay more and more attention to the research impact of applicants when deciding which researcher should be funded [2]. Research impact is also crucial for universities to decide which faculty applicant should be enrolled or which assistant professor should be offered a tenure. For a journal or conference, research impact, usually quantified based on citations [10, 18], is adopted to evaluate the venue quality and decide their rankings.

Because of the importance and popularity of impact evaluation, many quantitative impact metrics have been proposed based on the citation data to evaluate researchers, journals, and even research institutes. Citation count is one of the most straightforward and popular ways to view publication impact of a researcher, but it can be biased, since different disciplines or research topics may have greatly different citation patterns. h-index [18] is the most widely adopted impact metric, defined as the maximum number (say, h) of publications where each publication has at least h citations. However, it can also be misleading [25]. Suppose there are two researchers with the same h-index, but one has modest citations for almost all of his/her papers and the other has several highly cited papers but much fewer citations for the remaining papers. Clearly, such a difference can not be reflected by h-index. Another problem of h-index is the loss of temporal information, as it is unable to capture the impact difference between two researchers with the same h-index but significantly different *academic ages* (years since PhD study). A lot of other metrics based on h-index have also been proposed [19], for instance, h_f index and g index. However, these derived impact metrics still have their own drawbacks, and they are unable to show enough details of citation data to help users gain a deep insight into the research impact from different aspects.

Recently, researchers in the visualization and HCI fields have shown a great interest in visualizing the citation patterns of individuals [7, 26, 40]. However, these visualizations suffer from scalability issues or the lack of impact details. Other visualization research on publication citation focuses on visualizing either the global trend of a research field [16, 21] or the reference relationships between individual papers [27, 35], which, however, cannot be directly applied in the visualization of research impact of a researcher or a venue.

In this article, we aim at supporting better exploration and understanding of researcher impact. We propose an interactive and compact visual design based on matrix visualization, *ImpactVis*, to comprehensively show the details of research impact. It has better scalability and works for not only individual papers and researchers, but also venues. Since citation information is generally regarded as one of the most important factors for showing research impact, citation information is adopted as the main information for visualizing research impact. The major contributions of this paper can be summarized as follows:

- We thoroughly surveyed the previous work and conducted user interviews, from which the key factors for exploring the research impact through citation data have been identified.
- We proposed an interactive and compact visual design based on matrix visualization, which can intuitively show both the aggregated and detailed information of research impact, enabling better exploration and understanding of research impact for individual papers, researchers, and venues.
- We conducted case studies and in-depth expert interviews to demonstrate the effectiveness and usability of the proposed technique.

2 RELATED WORK

The related work of this article can be categorized into three parts: quantitative metrics of research impact, visual analysis of citation data, and matrix-based visualization.

2.1 Quantitative Metrics of Research Impact

Penfield et al. [28] systematically summarized the basic motivations of assessing research impact including evaluating the research value in order to inform funding decisions, demonstrating to government or tax payers the value of research and helping research organizations to understand and monitor their research performance. Because of all these reasons, the evaluation of research impact has become a hot research topic in recent years and a large number of quantitative impact metrics have been and continue to be proposed. Some researchers use citation count as an intuitive way to quantify research impact. While more researchers believe that research impact should incorporate both publications and citations. One representative quantitative metric for that is h-index [18], the most popular impact metric nowadays. However, since h-index is defined as the maximum number of papers h where each of them has at least h citations, it is subject to publication and citation distribution anomalies [25]. Egghe [10] introduced g-index, defined as the highest number of papers g that receive at least g^2 citations, in order to increase the sensibility of h-index to papers with high citations. To mitigate the bias when evaluating the impact of researchers across different scientific disciplines, h_s index [22] is introduced as the h-index normalized by the average h-index of all researchers in the same discipline. Similar ideas can also be found in the definition of h_f index [30] and Batista's $h_{i,norm}$ [3]. Stallings et al. [34] determines the scientific impact of a researcher by assigning relative credits to coauthors of each given paper and taking the academic age of a researcher into account, aiming at a fairer and more sensitive evaluation of researcher impact.

In summary, quantitative measurement of research impact has played such an important role in the academic community. Significant numbers of quantitative impact metrics have been proposed. However, almost everyone agrees that even the most sophisticated metrics are not able to fully capture the diversity and richness of research impact [14, 28].

2.2 Visual Analysis of Citation Data

Citations between different papers represent the academic relationship between papers and researchers. Therefore, an important goal of visualizing citation data is to show the relationship between papers and further provide an overview of research trends. For example, Citeology [27] shows the citation connection among a paper, the papers it references to, and the papers that cite this paper through visually linking them. PivotPaths [8] introduces an interactive visualization system for exploring multiple facets and relations such as authors, keywords, and citations. Zhao et al. proposed PivotSlice [44] to help users explore implicit and explicit relations of faceted datasets, especially publication and citation data. Stasko et al. present CiteVis [35], an interactive tool to portray the citation data of IEEE InfoVis conference. Heimerl et al. built CiteRivers [16] by clustering the paper contents of scientific literature collections and correlating papers with their citations, which offered an overview of the research trend in a specific research community. Lee et al. [21] introduced PaperLens to reveal research trends, connections, and activities in a conference through tightly coupled views. Wu et al. [40] made use of citation and publication data to visualize the career path of a researcher, using a line-chart-based design. Xie [41] visualized the citation patterns of computer science conferences. All these works can reflect the research impact of individual paper, researcher, or conference to some extent, but many details of research impact are lost and some drawbacks are inherent. For instance, Citeology shows the relationship between papers but may suffer from visual clutters, and CiteRiver is difficult to show the research impact of a paper or researcher.

Some related works are more focused on visualization of research impact. Shi et al. proposed VEGAS [31] to summarize the impact of an influential paper or individual researcher on different topics by showing a graph-based impact summarization, but it ignores the temporal information of research impact, which makes it unable to find the temporal changes of research impact. Maguire et al. [26] designed an impact glyph for each paper, showing the total citation counts and publication time of both the papers it references to and the papers that cite it. The glyph is further applied to visualizing the research impact of a researcher or institution. However, the visualization design suffers from serious visual clutter, even for a single highly cited paper, and is unable to well present the research impact of a prolific researcher. Kwon et al. [20] proposed *Scholar Plot* to visualize research impact of individual researchers. The basic design is a scatter plot, showing publication citation numbers along each year. However, more details, including how the research impact of individual papers evolves and what is the relationship between coauthors and citations, cannot be visualized by their designs. Cao et al. recently introduced Episogram [7] that can be applied to visualizing the temporal publication and citation patterns of a researcher, but it cannot show enough details of the citing papers and also suffers from visual clutter.

Different from previous work, *ImpactVis* aims at helping users comprehensively explore and understand the research impact in terms of different levels of details and temporal distribution. We also show that the techniques in this article have good scalability, which can be applied in impact visualization for individual papers, researchers, and conferences/journals.

2.3 Matrix-based Visualization

Matrix-based visualization has been widely applied in several types of data and different applications. Here, we focus on two types of data which are highly related to our work: graph data and sequential data.

According to the survey on graph visualization [38], matrix is one of the major visualization ways for showing static large graphs. For example, the large graph exploration tool named ZAME [11, 13] was proposed based on the combination of an adjacency matrix representation and automatic reordering algorithms. NodeTrix [17] used a hybrid approach to visualize social networks, where matrix visualization is the main part for supporting the analysis of graph communities. Lutton et al. presented an interactive matrix-based visualization tool for showing the interaction and information exchange between islands in the island-based parallel genetic algorithms [23], helping users characterize a good launch on the grid and tune the parameters of the algorithms. Zhao et al. [45] proposed a novel visualization technique based on matrix representation to analyze dynamic ego-networks. Blanch et al. [5] applied the matrix-visualization to the analysis of dendrograms and proposed a hybrid tree-matrix interactive visualization of dendrograms. Matrix-based visualization can also be combined with the traditional node-link diagrams to show graph data, e.g., visualizing semantic web data structures [1] and analyzing the co-clusters of bipartite graphs [42].

Matrix-based visualization is also popularly applied in the exploration of sequential data. GeneaQuilts [4] visualized the genealogy (i.e., the family relationships) as a diagonally-filled matrix with rows and columns representing individuals and nuclear families. Perer and Sun [29] visualized the clinical event sequences of patients as a temporal flow of matrices to unearth the hidden patterns within the clinical event sequences. Similarly, MatrixWave [46], a matrix-based visual design by concatenating the matrices with 45º rotation, is proposed to help users interactively analyze and compare event sequence data. EventAction [9], a visualization system to explore temporal sequence data, also used a matrix-based design for explaining recommendations of temporal sequences.

Inspired by the prior work on matrix-based visualization, we proposed an augmented matrix design to visualize the details of research impact that are unable to be easily conveyed by the existing methods.

3 ANALYTICAL TASKS

After surveying the previous work, we find that publication and citation are the most important and widely adopted proxies for indicating research impact. To better understand research impact and which aspects are important for visualizing research impact from the perspective of end users, we conducted informal interviews with three potential users. All of them are computer science researchers and two of them work on machine learning and deep learning, while the other user is focused on graph mining. During the interview, we asked the participants what they think are important for indicating the research impact and how they prefer to explore the research impact of a paper, a researcher, and a venue (conference or journal). Feedback from participants not only confirmed some analytical tasks introduced in previous work, but also offered us some new insights to the effective exploration of research impact. For instance, the users want to know the temporal details and influenced fields of a researcher's impact. They are also interested in the involvement of other researchers when exploring the research impact of an interested researcher. To accurately explain the concept in this article, we use the term *research followers* to refer to the researchers who cited the work of the interested researcher or venue. Since the visualizations of research impact of a journal/conference and a researcher share lots of similarities, we take the research impact of a researcher as an example and compile a list of analytical tasks as follows:

T1. Overview of research impact. How many citations has a researcher received? How many papers has the researcher published? How long has the researcher been working in the academic field (*academic age*)? What is the overall trend of self-citation of this researcher? Such kind of questions are important for gaining a quick understanding of the research impact of a researcher.

T2. General fields of research impact. What are the specific research fields of this researcher and which research fields are generally impacted by the research of this researcher? Since different research fields may have quite different publication and citation patterns, it is necessary to inform users of the researcher's general research fields and which fields his/her research impact goes to.

T3. Temporal evolution of publications and citations. How do the researcher's publications and citations evolve over time? In which year does the researcher publish most papers and in which year did his/her publications receive the most citations in total? For a specific paper or a specific year, how will the citations change over time? Research impact is essentially dynamic and answering these questions is helpful for gaining insights into the temporal patterns of research impact.

T4. Involvement of other researchers. Which coauthors of this researcher have more contribution to his/her publications and, further, which have a greater influence on his/her citations? In which years does a specific coauthor collaborate with this researcher? For the publications of a specific year, who are the major collaborators of this researcher? Who are the main research followers of this researcher and keep following and citing the work of this researcher? Answering these questions is helpful in understanding research impact from the perspectives of both research collaborators and followers. It is crucial for exploring the possible relationship between the research impact and coauthors and understanding on whom his/her research has an influence.

T5. Showing details on demand. Which paper of a researcher receives the most citations? What is the detailed information of a specific publication and all the papers that cite this specific paper? What are the impact differences between the most highly cited papers of each year? Users may want to know more details about a specific aspect of the research impact that they are interested in. Sometimes, they may want to filter the data; for example, one participant mentioned that he would like to focus on a researcher's publications and citations within the latest 5 to 10 years, especially when he wants to find a collaborator or follow the research work of an interested researcher. Therefore, it is important to enable user to interact with the visualization system and query on demand.

4 VISUAL DESIGN

In this section, we describe our design rationales and the final visual designs that we proposed for visualizing research impact.

4.1 Design Rationales

In response to the aforementioned tasks, we summarize the basic design rationales to guide our visual designs for showing research impact.

R1. Quick overview of research impact. According to the visual information seeking mantra, "overview first, zoom and filter, then details on demand" [32], it is suggested to provide users with an overview first and avoid making users overwhelmed by showing all the details at once, which is also consistent with our task analysis (***T1***). Considering that research impact is reflected in many aspects, a compact and clutter-free visual design is necessary for generating such a quick overview of research impact. In addition, it would be preferred if the visual design can support a preliminary impact comparison between different researchers or venues, since sometimes users may want to conduct a quick comparison between them.

R2. Comprehensive exploration of impact details. A comprehensive exploration of impact details can help users gain deep insights into the research impact of individual researchers and venues. Publications and citations are the most important two aspects that indicates research impact. Therefore, a visualization that shows the detailed information of publications and citations, including their general fields, quantitative amount, and temporal evolution, is key to the accurate understanding of research impact (***T2, T3***). It is also important to incorporate the collaborators of a researcher (or the authors of a venue) into the research impact analysis. Showing the correspondences between the publications and collaborators (or authors) can further reveal the potential reasons for the research impact (***T4***). Similarly, visualizing the details of research followers and the correspondences between citations and research followers are helpful for understanding which researchers are influenced by the publications of the interested researcher or venue (***T4***). Furthermore, to support an accurate understanding of impact details, it is helpful to provide users with the overall context in the visual designs when focusing on some impact details.

R3. Interactive exploration. Our informal interviews with potential users in the previous section showed that users may want to investigate the detailed aspects of research impact on demand (***T5***). Therefore, it is necessary to enable smooth interactions in the visualization system in order to help users better explore and understand the research impact details.

R4. Intuitive visual designs. Most users of the proposed technique would be general users who probably do not have a background of visualization. Intuitive designs would be more desirable than complicated ones, as such kind of designs are easier for users to understand and use, enhancing better exploration of research impact.

R5. Good scalability. Some previous designs [7, 26, 27] may suffer from visual clutter when there are a large number of publications or citations. Existing work of visualizing research impact may also be designed specifically for only a paper [27], a researcher [7], or a venue [41]. These two disadvantages limit the scalability of visualizing research impact. These undesirable issues should be well handled in *ImpactVis* to guarantee the good scalability.

4.2 Core Visualization Design

According to the analytical tasks and design rationales discussed above, we propose *ImpactVis*, an interactive visualization tool for better exploration and understanding of research impact for

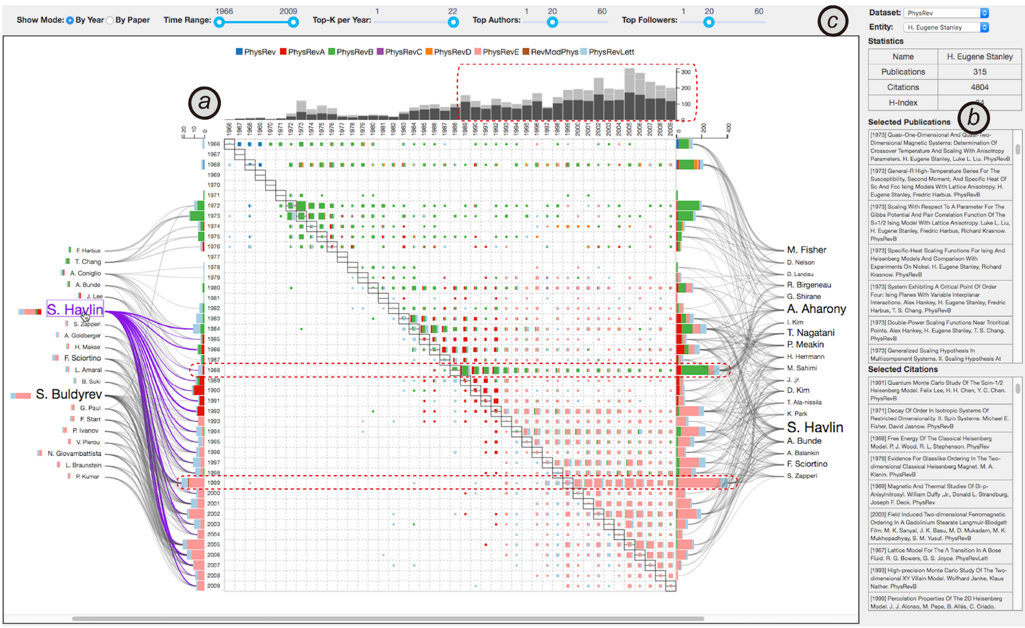

Fig. 1. User interface of *ImpactVis*, showing the research impact of Professor H. Eugene Stanley. (a) The impact view, which is also the main view of the user interface. The central part is a matrix in which each cell shows the citations or self-citations related to the current researcher. The stacked bars on the left and right side show the accumulated publications and citations in each year, respectively. The corresponding coauthors and research followers are shown and linked to their publications. (b) The tables showing the overall statistics about the researcher's research impact, publication, and citation list. (c) The control panel by which users can select the interested dataset and researcher, and filter the data. The dotted red rectangles highlight the findings using *ImpactVis*.

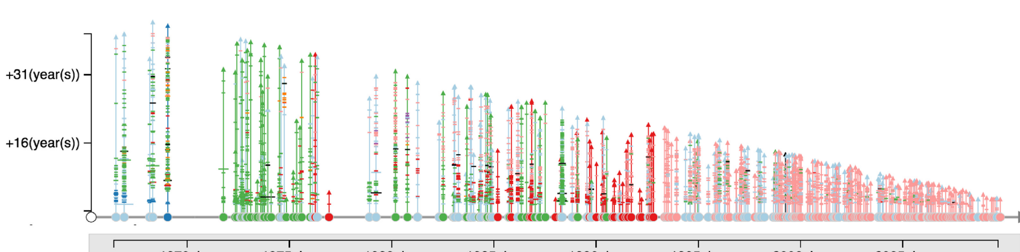

Fig. 2. The initiator view of Episogram [7] showing the research impact of Professor H. Eugene Stanley.

a paper, a researcher, or a venue. The user interface of *ImpactVis* contains three major parts (as shown in Figure 1): (a) the *impact view* based on matrix visualization, which is the main view of *ImpactVis* and shows the details of research impact; (b) the tables showing the overall statistics and detailed publication and citation information (***R1***); and (c) the control panel, in which users can select the interested researcher or venue, filter the publication and citation data, and adjust the main view (***R3***).

The aforementioned analytical tasks show that at least five aspects of information are important for visualizing and understanding the research impact of a researcher, i.e., publications, citations, self-citations, coauthors, and research followers. A naive design is to visualize the information using several separate views and then link these views, which is also our initial design for visualizing research impact. However, these five aspects of information are highly inter-connected, separated views make it difficult to gain a quick and deep understanding of such kind of connections between them. Taking into account that matrix is commonly used in many application scenarios and general users are familiar with it (***R4***), we chose a matrix-based visual design as the core visual design of *ImpactVis*, as shown in Figure 1(a). The impact details (e.g., publications, citations, self-citations, coauthors, and research followers) are clearly shown and their temporal distributions and inter-connections are also well expressed, making a compact visualization and supporting comprehensive explorations (***R2***).

The matrix is mainly used for visualizing publications and their citations or self-citations. Each row and each column represent one year. The stacked bars on the left side and right side show the numbers of publications and their total citations, respectively. The gray bars on top of a matrix represent the total citations a researcher received within each year and the unique citations are also indicated by black bars to provide users with a more accurate comprehension of the research impact in each year, as duplicate citations (i.e., one paper cites multiple papers of this researcher) may exist. Publications and citations may have different ranges, so independent axes are shown for the corresponding bars, informing users of the actual numbers of publications and citations.

The citations and self-citations are shown in the cells of the matrix where the details of their temporal distribution are clearly shown. The general citations are shown in the upper triangle of the matrix, as each paper can only be cited by those papers that are published no earlier than it. The cell at Row i Column j (where $j \geq i$) represents the citations that the papers published in Year i received in Year j. For the same reason, the self-citations will be shown in the lower triangle of the matrix, where the cell at Row i Column j (where $j \leq i$) shows the self-citations that the papers published in Year i cite his/her own papers published in Year j. When viewing one row as a whole, the stacked bars on the left and right show the total publications and citations, respectively, and the matrix cells on the left and right of this row show the temporal distribution of self-citations and general citations, respectively. Each row can also be changed to visualize the information of an individual paper through interaction.

Since it is important to show general fields of research impact instead of just simply showing the citation count, we categorize the citations and publications into different groups according to their research field and visualize them. Here, the groups can be determined by publication venues, general research areas, or different research topics, depending on the characteristics of a dataset and what users want to emphasize. Colors are naturally used to distinguish different groups. For the number of used colors, it is essentially determined by the number of research areas in the interested publication and citation data. When more colors are used, the information about research areas will be better delineated with more details, but more colors will bring more difficulty to users in understanding the visualization. Therefore, we need to strike a compromise. It is suggested the total number of colors used in a visualization system should not exceed "a dozen" [39]. In *ImpactVis*, we follow this rule by limiting the number of used colors to be less than "a dozen."

To effectively visualize the details about general field of research impact, we carefully designed the glyph within each matrix cell. A desirable glyph design should satisfy the following requirements: (1) being able to fully make use of the limited space within each cell, (2) clearly showing the overall (self)citation differences between cells, and (3) enabling an easy comparison between different types of (self)citations within and between cells. Our glyph design is shown in Figure 3(c). The total (self)citations within each cell are mapped to the area of the square, making it clear to see the overall citation difference between different cells. With fixed height, the number of each type of (self)citation is directly proportional to the width of the inner rectangles that form the whole

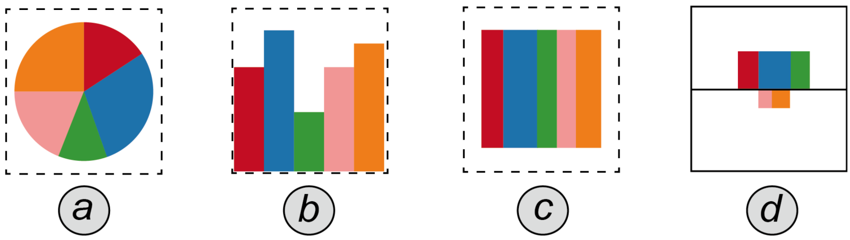

Fig. 3. Alternative glyph designs for showing (self)citations within each matrix cell. (a) Pie chart glyph. The area of the circle represents the total (self)citations and the inner angle encodes the relative percentage of each type of (self)citations. (b) Bar chart glyph. The height of each bar encodes each type of (self)citations. (c) Square glyph. The area of the square encodes the total (self)citations and each of the inner bars encodes one (self)citation type. (d) The glyphs at the diagonal cells encode both self-citation (lower part) and general citations (upper part) within one cell.

square. It makes it easy to compare different types of (self)citations. The (self)citation types are encoded using categorical colors. Before finally choosing the square glyph (Figure 3(c)), we also considered the pie chart glyph and bar chart glyph (Figure 3(a) and 3(b)), but discarded them at last. The pie chart glyph satisfies Requirement 1, but it does not help the comparison between the same type of (self)citations in different cells and is unable to fully make use of the cell space to encode the (self)citation difference between different cells, which is critical considering the highly limited space of each cell. The bar chart glyph matches Requirement 1 and 3, but makes the comparison of overall citations between cells difficult. The tricky part is the cells on the diagonal of the matrix, where both the self-citation and general citations should be visualized within the same cell, since a paper may not only cite but also be referenced by other papers published at the same year with that paper. This is solved by making use of the symmetry of square glyph, and showing the upper half square glyph of citation and the lower half square glyph of self-citation within a single cell.

The top-K coauthors and followers of a researcher are listed on the left and right side, respectively, where K can be interactively specified by users through the control panel (Figure 1(c)). The coauthors and followers are linked to the corresponding publication and citation bars in different years using smooth curves of which the width represents the relative number of publications and citations. The font size of the names of coauthors and followers encodes the number of publications or citations, and the vertical position of the name list is ordered according to the time of publications or citations. To further visualize the details of the collaboration with each coauthor, a stacked bar, where both the publication type and number are encoded, is shown beside each coauthor.

4.3 Interactions

ImpactVis supports rich interactions, which offer users a great capability to freely explore their interested aspects about the research impact of a researcher or a venue (***R3***). The most important interactions can be summarized as follows:

— ***Filtering.*** Filtering is one of the most important operations in *ImpactVis*. In the control panel of *ImpactVis* (Figure 1(c)), we provide the sliders to filter the interested time range, top-K highly-cited papers within each year, top (co)authors and top followers, which greatly helps users in the exploration of research impact. For example, in Figure 1, when a user wants to find out the top two highly cited papers in a specific period by Prof. Stanley, he/she can simply change the time range and specify the "Top-K per Year" slider as 2. The information of coauthors and followers will be updated simultaneously.

- ***Hovering and highlighting.*** Hovering is widely used for showing details of research impact in *ImpactVis*. For instance, when the user hovers on a coauthor, the coauthor and the links between the coauthor and the publication bars will be highlighted in purple (Figure 1), indicating the detailed years of collaboration. Similar operations are also enabled for the followers, helping users understand when and how the followers are influenced by the interested researcher. When the user hovers on a publication bar and citation bar, a tooltip will be shown, describing the number of publications or citations and their corresponding types of research field. When the user hovers on a circle representing a paper by the researcher, the paper information such as title, author, year, and venue will be shown in a tooltip. When the user moves the mouse on the matrix, the grid lines marking the row and column of the centered cell will be highlighted in red, helping the navigation on the matrix.
- ***Switching the detail level of publications.*** The details of yearly publications are shown by default. The user can switch to view all the details of each paper by simply changing the "Show Mode," or the user can zoom into paper details for only specific years by clicking the corresponding year labels, and the matrix will be expanded to show the self-citations and citations of all the papers of the interested years.
- ***Linking to the external paper URL.*** When the user finds some interesting papers and wants to further read these papers after exploration, he/she can simply click the circles representing each paper and *ImpactVis* will open a new page to show the detailed paper information using the DOI of each paper.
- ***Viewing (self)citations encoded in each matrix cell.*** Hovering on each matrix cell can inform users of the number and type of (self)citations encoded in this cell, but users may want to further know more details about the citation and referenced papers. Users can simply click the cell, then the publication and citation tables on the right (Figure 1) will be updated simultaneously, showing the title, author, published venue, and year of the corresponding papers.
- ***Showing a coauthor's detailed contribution to research impact.*** To help users better understand a coauthor's contribution to the research impact of a researcher or venue, the user can simply click the coauthor/author's name, then the information of all the papers that are collaborated by this coauthor is listed as matrix rows behind the row of each year (e.g., Figure 8), offering enough context to gain a quick impression about this coauthor's contribution in each year (***R2***).
- ***Finding coauthors/followers of a specific year.*** For a specific year, users may also want to know who actually contributes to the research impact of that year, or who are influenced by the research work of this researcher published in that year. Users can easily achieve it by simply clicking on the publication or citation bars of that year; then, all the corresponding coauthors or followers and the related links will be highlighted.

All the above interactions are easy to use, and they are extremely helpful and effective in the exploration of research impact, which will be discussed later in the subsequent expert interviews.

5 EVALUATION

To demonstrate the effectiveness of *ImpactVis* in helping the exploration of research impact, we first compared *ImpactVis* with the highly related existing work, Episogram [7], by using *Physical Review (PhysRev) Dataset*. Then, we further conducted in-depth expert interviews using *Microsoft Academic Graph (MAG) Dataset* [33] to provide evidence for the usefulness and usability of *ImpactVis*. The detailed findings and feedback from the actual users are reported.

5.1 Physical Review Dataset

Physical Review Dataset consists of the complete set of papers published by Physical Review and the citation relationships among them. The papers are from eight physics journals: *Physical Review (PR), Physical Review Letters (PRL), Reviews of Modern Physics (RMP), as well as Physical Review (PR) A, B, C, D, E.* Each of them focuses on a direction of physics. Thus, for each researcher, we can easily get all of his/her papers by querying the name. Then, we can further extract the citations of each paper and calculate the yearly citation number as well as the follower information. Like Episogram [7], we also take Professor H. Eugene Stanley, a famous American Physicist in statistical physics, as an example in order to better compare *ImpactVis* with Episogram and support the advantages of *ImpactVis*.

From Figure 1(a), it is easy to gain these findings:

(1) *Overall Trend of Research Impact:* Professor Stanley has been very productive for more than 40 years (1966-2009) and his research interests have shifted several times, which can be clearly observed from the color changes on the publication bars on the left side. In details, at the early stage of Professor Stanley's career (1971-1980), he mainly published his papers in PRB (green). Then he switched to PRA (red) from 1985 to 1992. While in the coming two decades (1993-2009), there is a burst of publications in PRE (pink). Taking into account that PRB specializes in condensed matter physics, PRA is mainly for atomic, molecular, and optical physics, and PRE for statistical physics and interdisciplinary physics, users can easily recognize the overall shift of his research interests. The similar changes are also observed in his citations (the citation bars on the right side of Figure1(a)), which offer an overall impression to where Professor Stanley's research impact goes (***R1***). Episogram (as shown in Figure 2) also provides such kind of an overview, but it is unable to provide users with an accurate perception on Professor Stanley‘s publications as well as citations in each year like *ImpactVis*.

(2) *Publication vs. Citation:* When viewing the publication bars and citation bars, it is obvious that the citations are generally received from the papers published in the same venue with Professor Stanley's publication. However, a large number of publications may not necessarily lead to high citations. For example, Professor Stanley published 22 papers and totally received 407 citations in 1999, which are the highest publications and citations. However, he published only six papers but received the second highest citations in 1988. This finding also implies that some papers in 1988 have significantly high research impact in terms of citations. All these detailed findings, however, are difficult to be gained from Episogram, confirming the advantage of *ImpactVis* in supporting comprehensive exploration of impact details (***R2***).

(3) *Detailed Citation and Self-citation:* By looking into the matrix cells, users can easily recognize the temporal patterns of citation and self-citation (***R2***). For example, it is easy to find that most of Professor Stanley's publications are still cited by other researchers after many years, indicating a long time range of research impact. It is also easy to find that from 1999 to 2009, Professor Stanley tends to more frequently cite his own previously written papers, indicating a stable and consistent research interest in statistical physics. However, it is unable to explore these details of research impact by Episogram.

(4) *Unique Citations:* The citation bars on top of the matrix show the citations that Professor Stanley received in each year. It is evident that his total citations in each year (gray bars) are increasing from 1989 to 2009, but the unique citations (the black bars) are generally stable. It informs users of the fact that other researchers tend to cite Professor Stanley's multiple papers at the same time which, however, cannot be achieved by Episogram.

(5) *Major Coauthors and Followers:* By using *ImpactVis*, users can immediately find out that Shlomo Havlin and Sergey Buldyrev are Professor Stanley's two major coauthors. By hovering

on their names, it is clear that Shlomo Havlin has been collaborating with Professor Stanley from 1984 to 2009 (Figure 1(a)). Similar explorations can be done about his followers, while Episogram does not support such kind of exploration.

In summary, *ImpactVis* provides users with a compact tool to gain a much deeper insight into a researcher's research impact than Episogram, demonstrating the ability of *ImpactVis* in visualizing research impact details (***R2***).

5.2 Microsoft Academic Graph Dataset

We also conducted in-depth expert interviews to observe how the actual users would use *ImpactVis* in practice and collect their feedback. The *Microsoft Academic Graph (MAG) Dataset* [33], which contains scientific publication records and the citation relationship between them, is used for our expert interviews. We chose MAG dataset for the study instead of Physical Review Dataset or other public citation datasets (e.g., Aminer [36]) because of two reasons: one is that the papers and citations of MAG dataset are from a much larger number of venues than *Physical Review Dataset* and it can better reflect the real research impact of researchers and venues. The other is the completeness of citations. Before finally choosing MAG dataset, we quickly investigated several public citation datasets and found that MAG dataset preserves relatively more complete citations than the other datasets, though some publication and citation data are also lost.

5.2.1 Data Preparation. For the MAG dataset, each publication has a unique ID and its detailed information (e.g., title, authors, published venue, and time) is also preserved. The citation relationships between papers are stored in a citation table, where the ID of a paper and an ID of its citation paper are written in a row. By making use of this information, we can easily process the raw data to explore the research impact of interested researchers and venues, which is similar to the process of Physical Review Dataset. We extracted the publications and citations of a set of researchers including some leading scholars in the field of visualization, human-computer interaction (HCI), data mining, and computer vision. The publications and citations of a list of top conferences and journals in the same field are also processed, such as TVCG, CHI, CSCW, CVPR, and ECCV. To show the general research fields of individual papers, researchers, and venues, we follow the categorization by China Computer Federation (CCF) [12] and categorize all the related journal/conferences in the computer science field into 10 groups: *system and architecture; computer network; security; software engineering; database and mining; theory; graphics, visualization, and multimedia; artificial intelligence (AI), machine learning (ML), and computer vision (CV); HCI; and interdisciplinary.* Since not every venue can be explicitly categorized into one of the 10 groups, we added an extra type labeled as "Others." Such categorization is manually labeled by one author of this article, since there are no good automated methods that can categorize all those venues perfectly. To speed up the manual labeling, we also adopted non-negative matrix factorization [43] based on the paper keywords in the MAG dataset to cluster venues, which helps us filter venues that do not belong to computer science and quickly label a group of similar venues.

5.2.2 Participants. We invited 10 expert users to participate our in-depth expert interviews. These expert users are either postgraduate students from the computer science department of a university or visiting scholars in an industrial research lab. All of them have at least 3 years' experience in doing research and are not involved in this project. Their main research interest ranges from data visualization, computer vision, machine learning, and human-computer interaction. Almost all of them have published at least two research papers in their corresponding research fields. The ages of the participants range from 23 to 29 and two of them are female. No gift or compensation is offered to the participants. Their detailed demographic profiles are summarized in Table 1.

Table 1. The Personal Information of Expert Users Participated in the In-depth Expert Interviews

User	Age	Sex	Research Interest
1	29	M	visualization, computer vision
2	27	M	visualization, HCI
3	27	M	graph mining, machine learning
4	26	M	visualization
5	24	M	HCI
6	27	M	visualization
7	26	M	HCI, visualization
8	23	M	computer vision
9	23	F	visualization
10	26	F	computer vision, HCI

Table 2. The Post-study Questions. Q1-4 Are Evaluating the Overall Usability of *ImpactVis*. Q5-8 are Assessing the Effectiveness in Revealing Impact Details

Q1	Is it easy to understand the overall visual design?
Q2	Is it easy to find out the overall publication and citation trend?
Q3	Is it easy to know the major coauthors and research followers?
Q4	Is it easy to find out the major research fields and impacted fields?
Q5	Is it easy to find out the correspondence between publications and their citations?
Q6	Is it easy to find out the detailed yearly (self-)citations of each paper or yearly publications?
Q7	For a coauthor (or research follower), is it easy to recognize the time of collaboration (or citation)?
Q8	For the publication(s) of a year, is it easy to find out their (co)authors and research followers?

5.2.3 Procedures. The in-depth expert interviews were conducted with each participant one by one. When interviewing each expert user, we first briefly explained the motivation of our project and introduced the usage of the whole user interface and basic operations of *ImpactVis* to them. Then, we asked the users to freely explore the publication and citation data of both individual researchers and venues that they are interested in, which will make the expert users familiar with the basic operations of *ImpactVis*. After that, several specific leading researchers (e.g., Jeffrey Heer and Tamara Munzner who are established researchers in visualization and HCI) and top conferences were required to explore in order to check the consistency of observations on research impact and provide evidence on the effectiveness of *ImpactVis*. To gain detailed feedback from users, a thinking aloud method was used and we asked users to comment verbally during their exploration. We kept observing the process of each user and wrote down their major operations, findings, and comments. After the users finished the exploration, we asked them several analytical questions about the research impact of a researcher or venue in order to confirm whether the user can easily use *ImpactVis* to gain insights into the research impact. Finally, we collected users' general comments about *ImpactVis* including the possible limitations and improvement suggestions. To gain a more accurate understanding about the usability of *ImpactVis*, we also designed a post-study questionnaire, where users responded to eight questions on a 5-point Likert scale (1=very easy, 5=very difficult). The former four questions focus on evaluating the overall usability of *ImpactVis* and the latter four questions mainly assess its usability in revealing impact details, as shown in Table 2.

The expert interviews were conducted on a MAC laptop (15-inch screen, 16GB RAM, 2.5GHz CPU), and by the consent of users, we also used an audio-recording software to record all the verbal comments of the users for further analysis. The whole interview with each user usually took between 60 and 100 minutes.

5.2.4 User Exploration and Findings. During the interviews, we have recorded the whole exploring processes. In this section, we summarize how users interact with *ImpactVis*, and detailed findings are also reported.

Overview

The exploring process started with loading the citation data of a researcher or a research venue. In most cases, the participants chose researchers they were familiar with for the initial exploration. For example, some users (e.g., User 1, User 4, User 6, and User 10) selected their PhD advisor(s); other users may choose their collaborators or famous researchers in their fields. Since *ImpactVis* provides the rich context of the citation data in a static view, during the interview, the participants spent more time in exploring and describing the information they got from *Impact View*. The knowledge they acquired by their exploration order are listed as follows:

— *Research Area:* The research areas of publications and citations are encoded as the color of bar charts and matrix cells, which can easily catch users' attention. The participants, in most cases, began with describing the research area the researchers work on as the starting point. For instance, after loading the data of his advisor, User 1 immediately gained the impression that both the publications and citations of his advisor are mainly from the visualization field according to the color encoding scheme, which makes sense to him. User 2 also explored the research impact of several venues including some conferences that do not belong to his own research field. For example, when he began to explore the impact of the conference ECCV, a top conference in computer vision, he quickly gained an initial impression that ECCV has an impact mainly on AI/ML/CV and graphics/visualization/multimedia, indicated by the yellow and blue color in Figure 6 (***R1***). Both User 7 and User 8 explored the research impact of their common coauthor. Before using *ImpactVis*, they only know that their coauthor has a research interest in HCI and visualization. But by using *ImpactVis*, both are surprised to find that their common coauthor also has many papers in the data mining field, in fact. When User 10, who recently switched her research interest from computer vision to visualization, was studying Jeffrey Heer's research impact, she said it is very easy for her to figure out that Prof. Heer is interested in both visualization and HCI (***R1***). The impacted research areas and the details about which year and which fields the citations are from are also clearly visualized in *Impact View*.

— *Collaborators and Followers:* The name lists of collaborators and followers are symmetrically located on the left and right side of *Impact View*, respectively. The name lists and the links between the names and the central matrix can give users a quick understanding about a researcher's collaboration with others and people influenced by this researcher. In our in-depth user interviews, all the expert users said that it was very easy for them to figure out who were the major coauthors and research followers, as the names of coauthors or research followers who closely collaborate with or follow the interested researcher more frequently have bigger sizes. The corresponding collaboration and following time period can also be clearly shown by simply hovering on the names (***R2***). For instance, User 4 investigated his advisor's research impact and found that all the major coauthors of his advisor are the former PhD students of his advisor. User 5 explored a young researcher whose research interests are visualization and HCI, and was surprised to find that one of his major coauthors is from the field of data mining. User 5 further checked it with *ImpactVis* and found that they had coauthored a few papers combining data mining and visualization techniques.

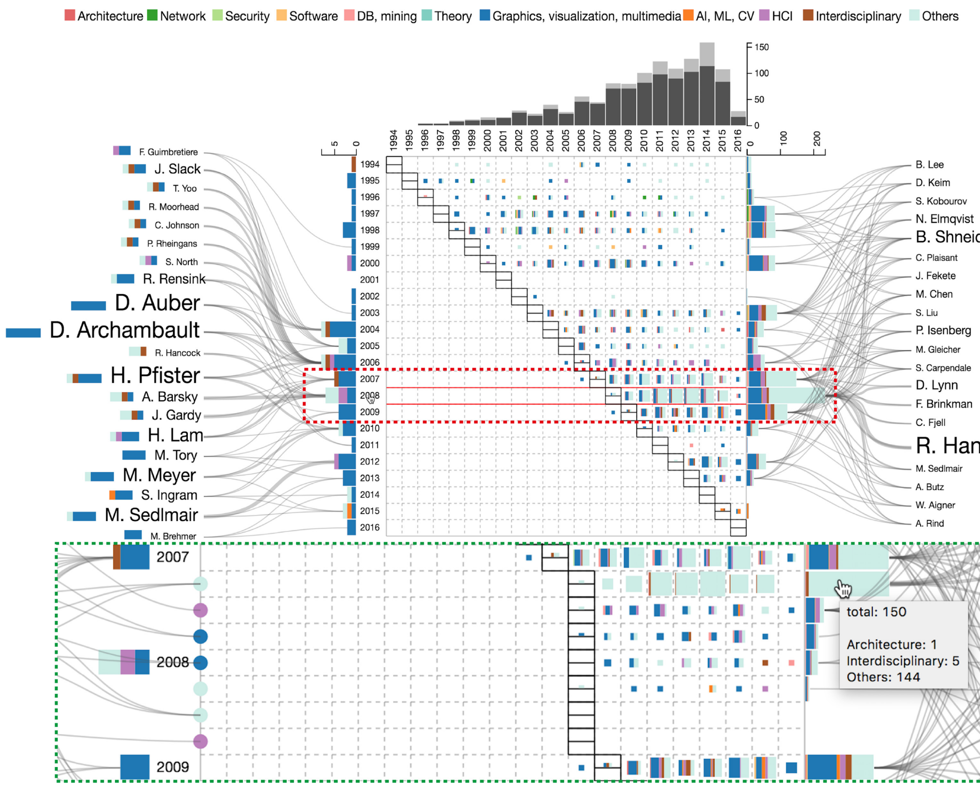

Fig. 4. The main view of *ImpactVis* showing the details about research impact of Tamara Munzner. The region marked in green rectangle corresponds to the same region of the red rectangle when the user clicked the year label in order to expand the publications in 2008 and see the details of each paper. The user easily noticed that the citations of the most cited paper in 2008 are mainly from the field type "Others".

Before using *ImpactVis*, User 2 believed that the coauthors of a researcher have a high probability of being the followers. However, contrary to his initial expectation, the followers of Professor Munzner and Heer are quite different from their coauthors. Among their top 20 followers, many leading researchers in visualization and the HCI field are noticed, such as Ben Shneiderman, Bongshin Lee, Niklas Elmqvist, and Jean-Daniel Fekete (as shown in Figures 4 and 5). User 2 said that it can be treated as an indication of Professor Munzner and Heer's significant research impact in visualization and the HCI field, since their work is widely recognized by other established researchers.

Several participants (e.g., Users 2, 3, 5, and 7) also agree that *ImpactVis* can help in finding new interested researchers and exploring new research field. Users 3, 5, and 7 further explored the research impact of top coauthors or research followers after they investigated a selected researcher. They said that if they were interested in a researcher, they probably also had strong interest in the coauthors who closely collaborate with this researcher. *ImpactVis* definitely offers them a convenient way to do that. After exploring the research impact of ECCV, User 2 also commented that *ImpactVis* may be quite useful if he wanted to start doing some research about computer vision, since he could easily know who are the leading researchers by viewing the top-author lists on the left. For example, he immediately

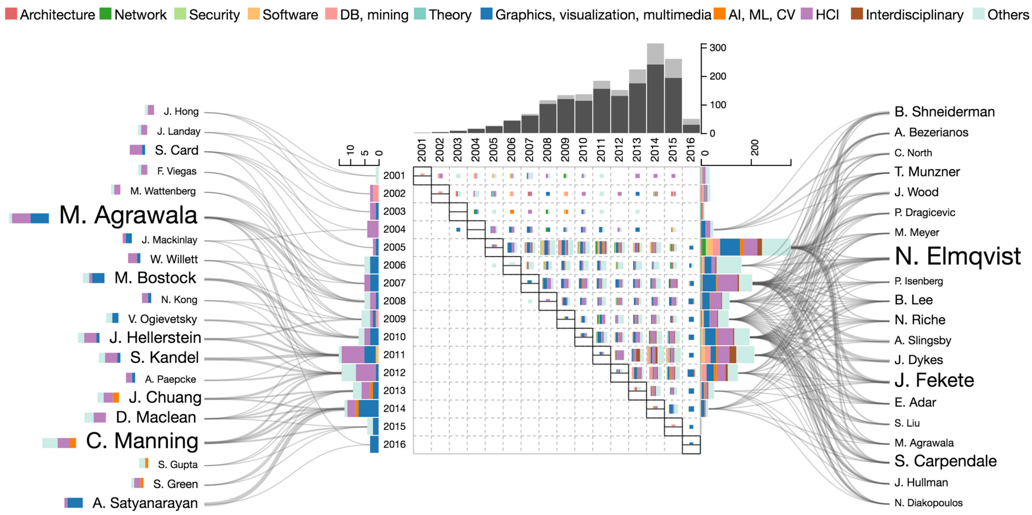

Fig. 5. The main view of *ImpactVis* showing the details of the research impact of Jeffrey Heer.

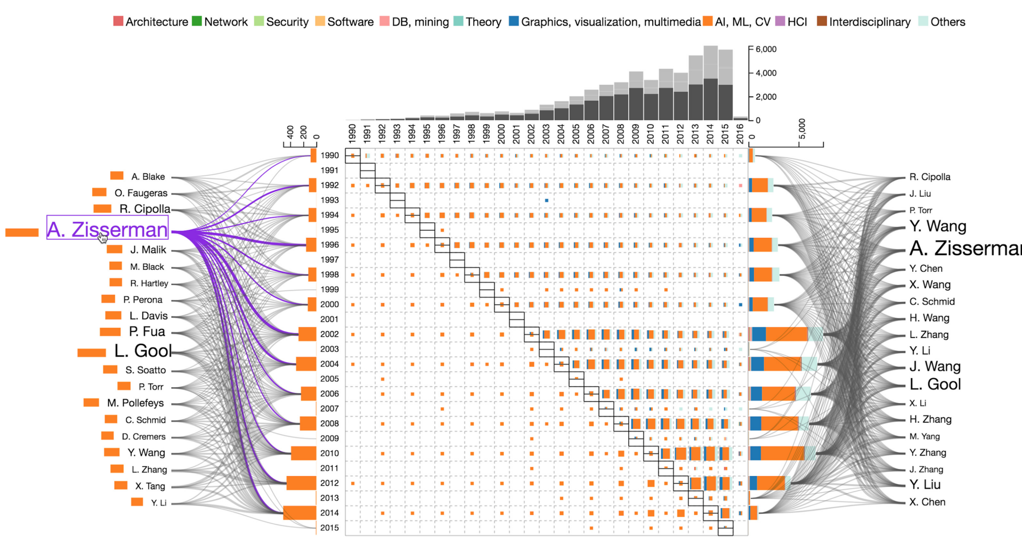

Fig. 6. The main view of *ImpactVis* showing the details of research impact of ECCV. The author Andrew Zisserman and the links to the corresponding publication bars are highlighted when the user hovers over it.

knew that Andrew Zisserman seemed to be a leading researcher in computer vision since 1994 through hovering on his name, as shown in Figure 6. He could also filter the top-cited papers in each year and start by reading those representative papers.

User 8 also noticed that Professor Munzner does not have any stable collaborators in the early stage (1994-2003) of her academic career (Figure 4), while Professor Heer has consistently collaborated with several researchers in the first ten years (2001-2010) of his academic career (Figure 5). User 8 said that such a difference may indicate two different ways of growth in one's academic career.

— *Temporal Information:* The overall temporal information about research impact of a researcher or venue is another major type of finding when participants were using *ImpactVis* (***R1***). For example, with *ImpactVis*, users could easily recognize a researcher's *academic age* (i.e., the time range that a researcher is involved in academic research), the overall trend of publications and citations, the years with more publications or citations, and the time range that a coauthor collaborates with the interested researcher.

More specifically, User 10 easily recognized that Professor Heer's academic career starts around 2001 and has been very active in the past 16 years, especially the years between 2011 and 2014 where he published more than 13 papers every year, as shown in Figure 5. By hovering on the name tag "M. Agrawala," the coauthor who has most collaborations with Professor Heer, User 9 quickly knew the major collaboration time between them is 2006-2012, as shown in Figure 5.

One interesting finding about Professor Munzner from User 1 is that he noticed Professor Munzner's publications in 2008 received the most citations, but the majority of citations are from the field type "Others" (Figure 4). At first glance, he conjectured that there could be some errors in the data, but when he clicked the year label "2008" to expand and inspect the details of each paper, he realized that all those "suspicious" citations were due to the paper InnateDB [24], which is Professor Munzner's collaborated work with researchers from biology and medical science, and he received lots of citations from the medical and biology fields (categorized as "Others" in *ImpactVis*). Before using *ImpactVis*, User 1 said he did not notice that Professor Munzner had such a paper.

User 1 also noticed that many of the self-citations of researchers are lost, which is due to the missing data in the MAG dataset. When exploring the research impact of the venues, he noted that *ImpactVis* is not as smooth as exploring the research impact of individual researchers because of the larger number of publications and citations of venues. Similar to User 1, User 2 also first randomly selected some leading researchers in his interested field (visualization) and freely tried the interactions of *ImpactVis* that we introduced.

As a layman of the computer vision field, User 2 noticed that the publications and citations are missing in some years and asked us if there is any problem in the data, but finally understood the reason after we reminded him that ECCV is held every two years. He (User 2) also noticed an obvious increasing trend in both publications and citations of ECCV, which is clearly indicated by the stacked bars on the left, right, and top of the center matrix in Figure 6. He said it may be a sign of increasing impact of this conference and he really enjoyed the good scalability of *ImpactVis* in supporting the impact exploration of both individual researchers and individual venues (***R5***).

Filtering on Demand

The rich interactions in *ImpactVis* make it convenient for users to filter the publication and citation information and gain a deep insight into the research impact (***R3***). For example, User 1 is very interested in exploring the highly cited papers of a researcher or venue. He said that this is one of his favorite functions when using *ImpactVis*, because it is usually very important to get to know the most representative work of a researcher or venue and *ImpactVis* makes such kind of exploration very easy and intuitive. By sliding the "Top-K per Year" slider and clicking the "Show Mode" checkbox as "By Paper," he was able to clearly view the temporal evolution of citations of all the representative papers and conveniently link to the DOI page of each paper when he wanted to see original articles.

When User 2 did a filtering to view the highly-cited work of Professor Heer, he immediately noticed the most highly cited paper Prefuse [15], in which Professor Heer introduced the widely used

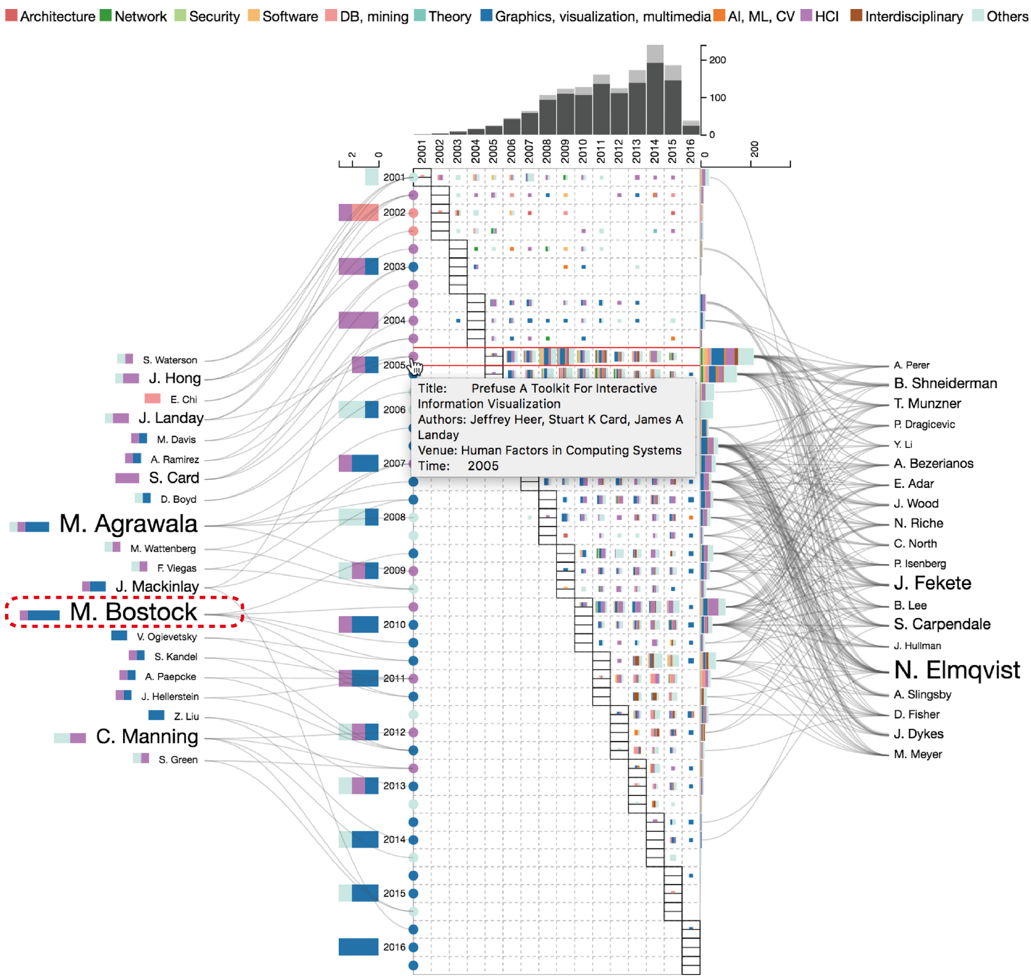

Fig. 7. The research impact of Jeffrey Heer after the user did filtering and kept only the top three highly-cited papers in each year. The matrix view is shown in the mode "By Paper" and the user easily observed the highly cited paper Prefuse [15] and noticed that "Mike Bostock" pops out in terms of size in the coauthor list.

software tool for creating data visualizations. Another interesting finding by User 2 after filtering is that Mike Bostock, who is Professor Heer's student and developed the dominated visualization tool D3 [6], pops out by the size in the coauthor name list (as shown in Figure 7). It implies that their collaboration generally leads to high-impact research and Mike Bostock contributes a lot to the research impact of Professor Heer. To gain a more accurate understanding about Mike Bostock's contribution to the research impact of Professor Heer, User 2 clicked the coauthor name to view more details, as shown in Figure 8.

Other users also had several such kind of similar findings when using *ImpactVis*.

Visual Comparison

When participants were exploring the citation from researchers with similar backgrounds, we noticed sometimes they switched the data to compare the research impact of different researchers. For example, after exploring the cases of Tamara Munzner (TM) and Jeffrey Heer (JH), User 1 noticed

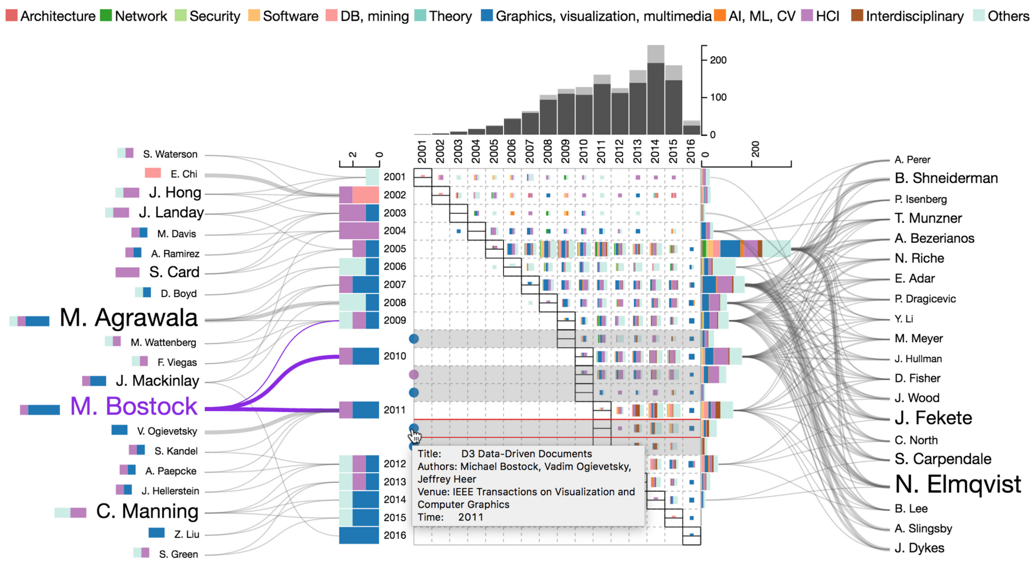

Fig. 8. The main view showing the research impact of Jeffrey Heer where only the top three highly cited papers in each year are kept. After the user clicked the coauthor name "Mike Bostock," all the corresponding papers coauthored with Mike Bostock are appended behind each year (the gray rows), enabling a better understanding of this coauthor's contribution to the total research impact.

that though they have a similar *h*-index (TM: 20, JH: 22), the details of their research impact are quite different when comparing Figures 4 and 5—Professor Munzner has a relatively long *academic age* (1994-2016) and her publications and citations are dominated by visualization, while Professor Heer started his academic career several years later (2001-2016) and is interested in both visualization and HCI. User 9 further recognized that though Professor Heer has a relatively shorter academic age than Professor Munzner, he has published relatively more papers and received more citations. When taking into consideration only the recent years (2011-2016), Professor Heer is also more active in terms of publications and the received citations each year.

5.2.5 Usability Evaluation. As mentioned above, after each user finished exploring the research impact of individual researchers and venues, we first asked them several analytical questions about the research impact of individual researchers or venues. We found that all the expert users can easily and accurately find the answers with the help of *ImpactVis*. We also asked them to assess the usability of *ImpactVis* from the perspectives of exploring both overall and detailed research impact, based on Table 2. The result is shown in Figure 9.

The results show that *ImpactVis* has an excellent usability in helping users explore both the overall and detailed research impact of individual researchers or venues. Almost all the users agreed that *ImpactVis* makes the exploration of both the overall and detailed research impact either *easy* or *very easy* (***R4***), though some of the expert users actually do not have any background in visualization or human-computer interaction. No one reported that *ImpactVis* is *difficult* or *very difficult* to use and there is only one user rated Q2 and Q7 as "3" (*"neither easy nor difficult"*), and only two users believed that Q5 also belongs to "3." These user studies indicate that *impactVis* has very good usability and makes the exploration of research impact of individual researchers or venues pretty easy. When comparing the difference between exploring the overall research impact (Q1-4) and detailed research impact (Q5-8), it is obvious that users generally believe that exploring the

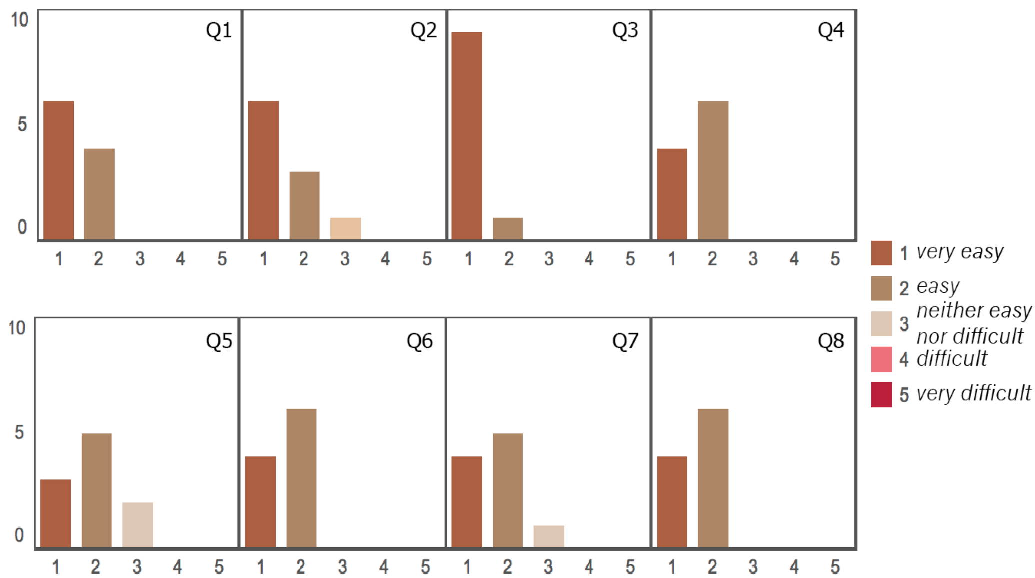

Fig. 9. The result of the post-study questionnaire showing all the expert users' responses to the eight questions on a 5-point Likert scale.

overall research impact is relatively easier than the detailed research impact. Three users directly pointed out that one possible reason for this is that the overall information about the research impact is directly shown without any interactions, while interactions may be more often needed in exploring the details of the research impact.

In summary, the post-study questionnaire indicates that *ImpactVis* has good usability. It is easy to use and can effectively help users explore the overall and detailed research impact of individual researchers or venues.

5.2.6 Overall Comments on ImpactVis. We further asked them for their overall comment on *ImpactVis* including possible flaws. Their feedback is summarized as follows:

Intuitive and Compact Visual Design. All the users liked the visual design of *ImpactVis*. Like what is reflected in Figure 9, the visual design is intuitive and easy to understand because all the users, including Users 3, 8, and 10 without a background in visualization, can easily and quickly understand the usage of *ImpactVis* (***R4***). Users 8 and 10 also explicitly pointed out that the symmetrical arrangements of publications and citations, coauthors and research followers make the whole visualization much better for understanding. The visual design of *ImpactVis* is compact. User 2 explicitly commented that by visualizing the general citations at the upper triangle and self-citations at the lower triangle of the matrix, it not only shows sufficient temporal details about research impact but also fully makes use of the matrix space. By packing the publication and citations bars on the left, right, and top of the matrix, it is quite easy to understand the details of publication and citation from different perspectives, enabling deep comprehension of research impact. Other expert users had similar comments with User 2 and also believed that the square glyph within each matrix cell fully makes use of the limited space within each cell.

Straightforward Interactions and Great Usability. Simple and straightforward interactions are another aspect that all the users enjoyed most. These interactions significantly facilitate the exploration of research impact (***R3***). Owing to the intuitive visual designs and interactions, *ImpactVis* supports exploring the research impact of both individual researchers and venues easily, demonstrating its good scalability and great usability.

Limitations and Improvement Suggestions. Users also pointed out some limitations they found when using *ImpactVis*. For example, the relatively small size of the matrix cell is the issue mentioned by User 1, since the small size sometimes makes it difficult to distinguish the different citation types within each cell. He suggested choosing a large fixed size for each matrix cell or using an adaptive size based on the number of matrix rows and columns. Considering that an adaptive cell size may be misleading (e.g., the same amount of citations may be represented in different cell sizes) and is not suitable for comparison among researchers or venues, we did not choose it in the final design. For using a fixed large size for matrix cells, we need to strike a compromise, as a large cell size may make it unable to show a quick impact overview for some researchers with many matrix rows (long academic ages) within the limited screen, which is undesirable. Taking all these factors into consideration, we only slightly increased our matrix cell size, ensuring that the increased cell size is suitable for the common cases in the dataset. In addition, we also added an interactive tooltip for each matrix cell to show the citation details. When users hover over a matrix cell, a tooltip recording the citation details such as citation types and amount will be immediately shown, mitigating the disadvantages resulted from the relatively small cell size. Users 3 and 6 mentioned that the interactions of exploring venue impact are not as smooth as that of exploring the impact of individual researchers, since the venue data usually has large data sizes, making it a bit slow to process the data and render the visualization. It has been improved through optimizing the implementation of dataset preprocessing and visual rendering, enabling more smooth interactions in the exploration of venue impact.

6 DISCUSSION

Based on the above detailed case study and in-depth expert interviews, we demonstrate the significant advantages of *ImpactVis* in visualizing the research impact of individual papers, researchers, and venues. However, there are still three major aspects that need further detailed discussions.

Factors Indicating Research Impact. In *ImpactVis*, we mainly visualize the research impact by using the publication and citation data, as they are popularly used by many researchers on research impact [3, 10, 18, 30]. However, there is still not a universal consensus on the definition of research impact [2, 37]. For example, some people argue that citations cannot accurately reflect the impact of a paper, since not all the readers of a paper will necessarily publish another paper and cite the paper he/she reads. Because of this, Ball and Duke [2] suggested that the number of page views, downloads, and social media links (and the like) may be more appropriate for evaluating research impact. Others would argue that not all the research work will end up in a paper. For taking all these aspects of factors into account, we may leave it as future work.

Citation Data. The citation data used in this article is from publicly available datasets. The data itself has some noise and also loses some citations, which may not strictly reflect the actual research impact of a paper, researcher, or venue. However, as the focus of *ImpactVis* is on the visual design technique, *ImpactVis* can be directly applied on the other citation dataset, when the dataset with better quality is available.

Classification of Publications and Citations. In *ImpactVis*, we categorize all the papers and citations into different types based on the manually labeled types of their venues, in order to reflect their basic research field. It may not always be the optimal choice, as it only reflects the high-level research field. To gain a more detailed classification of the publication and citation types, one can also choose some other methods (e.g., use topic modeling when detailed text data of each paper is available) to classify the papers as long as it satisfies the needs of users.

Scalability. We have demonstrated that *ImpactVis* has good scalability, since it can effectively support the impact analysis of individual papers, researchers, and venues, helping users understand both the impact overview and details. However, another scalability issue is related to the data size (e.g., the rows and columns that the matrix contains). When a researcher has a long academic age or a large number of publications, *ImpactVis* may not be able to show all the matrix rows/columns simultaneously within the limited screen. To handle this issue, *ImpactVis* enables users to interactively filter interested subsets and shows details on demand.

7 CONCLUSION

Previous work on quantifying or visualizing research impact is not able to provide enough impact details or suffers from visual clutter and scalability issues. In this article, we first identified the key factors for visualizing research impact through surveying previous work and organizing user interviews, and a list of essential analytical tasks and design rationales were summarized. Guided by these tasks and design rationales, we proposed a compact and intuitive visualization technique, *ImpactVis*, to help users explore the research impact of individual papers, researchers, and venues. A set of straightforward and quite useful interactions were also added to *ImpactVis*.

Then, the proposed technique was comprehensively evaluated. We not only conducted a detailed case study but also organized in-depth expert interviews with participants from different research fields. The case study compared *ImpactVis* with a representative previous work using exactly the same data, demonstrating the capability of *ImpactVis* in delineating much more detail than the existing work. The in-depth expert interviews were based on a publicly available dataset. Many interesting findings in this real dataset were found by the expert users, providing substantial support for the good scalability and effectiveness of *ImpactVis* in visualizing research impact. The post-study questionnaire further demonstrates the excellent usability of *ImpactVis*. Its possible limitations are also discussed. In future work, we would like to conduct more user studies to further demonstrate the effectiveness of *ImpactVis* and consider more broadly-related factors (e.g., patents, research talks, and industry products) to comprehensively analyze research impact. Also, we plan to extend the proposed technique to analyze the impact of a post or user in social media like Twitter.

ACKNOWLEDGMENTS

The authors would like to thank the expert users for participating our in-depth user interviews, P. Deville and R. Sinatra for providing us with the Physical Review Dataset, and Microsoft Research for releasing the Microsoft Academic Graph Dataset that is used in this article.

REFERENCES

[1] Benjamin Bach, Emmanuel Pietriga, and Ilaria Liccardi. 2013. Visualizing populated ontologies with OntoTrix. *International Journal on Semantic Web and Information Systems* 9, 4 (Oct. 2013), 17–40. DOI: http://dx.doi.org/10.4018/ijswis.2013100102

[2] Alex Ball and Monica Duke. 2015. How to track the impact of research data with metrics. *DCC How-to Guides* (2015). http://www.dcc.ac.uk/resources/how-guides/track-data-impact-metrics.

[3] Pablo D. Batista, Mônica G. Campiteli, and Osame Kinouchi. 2006. Is it possible to compare researchers with different scientific interests? *Scientometrics* 68, 1 (2006), 179–189.

[4] Anastasia Bezerianos, Pierre Dragicevic, Jean-Daniel Fekete, Juhee Bae, and Ben Watson. 2010. GeneaQuilts: A system for exploring large genealogies. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (Nov. 2010), 1073–1081. DOI: http://dx.doi.org/10.1109/TVCG.2010.159

[5] Renaud Blanch, Rémy Dautriche, and Gilles Bisson. 2015. Dendrogramix: A hybrid tree-matrix visualization technique to support interactive exploration of dendrograms. In *Proceedings of the 2015 IEEE Pacific Visualization Symposium (PacificVis)*. 31–38. DOI: http://dx.doi.org/10.1109/PACIFICVIS.2015.7156353

[6] M. Bostock, V. Ogievetsky, and J. Heer. 2011. D3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (Dec. 2011), 2301–2309. DOI: http://dx.doi.org/10.1109/TVCG.2011.185

[7] Nan Cao, Yu-Ru Lin, Fan Du, and Dashun Wang. 2016. Episogram: Visual summarization of egocentric social interactions. *IEEE Computer Graphics and Applications* 36, 5 (Sept. 2016), 72–81. DOI: http://dx.doi.org/10.1109/MCG.2015.73

[8] Marian Dörk, Nathalie Henry Riche, Gonzalo Ramos, and Susan Dumais. 2012. PivotPaths: Strolling through faceted information spaces. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (Dec. 2012), 2709–2718. DOI: http://dx.doi.org/10.1109/TVCG.2012.252

[9] Fan Du, Catherine Plaisant, Neil Spring, and Ben Shneiderman. 2016. EventAction: Visual analytics for temporal event sequence recommendation. In *2016 IEEE Conference on Visual Analytics Science and Technology (VAST)*. 61–70. DOI: http://dx.doi.org/10.1109/VAST.2016.7883512

[10] Leo Egghe. 2006. Theory and practice of the g-index. *Scientometrics* 69, 1 (2006), 131–152.

[11] Niklas Elmqvist, Thanh-Nghi Do, Howard Goodell, Nathalie Henry, and Jean-Daniel Fekete. 2008. ZAME: Interactive large-scale graph visualization. In *2008 IEEE Pacific Visualization Symposium*. 215–222. DOI: http://dx.doi.org/10.1109/PACIFICVIS.2008.4475479

[12] China Computer Federation. 2005. A list of computer science conferences and journals recommended by CCF. (2005). Retrieved Oct. 27, 2016 from http://history.ccf.org.cn/sites/ccf/paiming.jsp.

[13] Jean-Daniel Fekete, Niklas Elmqvist, Thanh-Nghi Do, Howard Goodell, and Nathalie Henry. 2007. *Navigating Wikipedia with the Zoomable Adjacency Matrix Explorer*. Research Report RR-6163. INRIA. 25 pages. https://hal.inria.fr/inria-00141168.

[14] Jane Grimson. 2014. Measuring research impact: Not everything that can be counted counts, and not everything that counts can be counted. *Bibliometrics. Use and Abuse in the Review of Research Performance* (2014), 29–41.

[15] Jeffrey Heer, Stuart K. Card, and James A. Landay. 2005. Prefuse: A toolkit for interactive information visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'05)*. ACM, New York, NY, 421–430. DOI: http://dx.doi.org/10.1145/1054972.1055031

[16] Florian Heimerl, Qi Han, Steffen Koch, and Thomas Ertl. 2016. CiteRivers: Visual analytics of citation patterns. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (Jan. 2016), 190–199. DOI: http://dx.doi.org/10.1109/TVCG.2015.2467621

[17] Nathalie Henry, Jean-Daniel Fekete, and Michael J. McGuffin. 2007. NodeTrix: A hybrid visualization of social networks. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (Nov. 2007), 1302–1309. DOI: http://dx.doi.org/10.1109/TVCG.2007.70582

[18] Jorge E. Hirsch. 2005. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America* 102, 46 (2005), 16569–16572. DOI: http://dx.doi.org/10.1073/pnas.0507655102

[19] Jasleen Kaur, Filippo Radicchi, and Filippo Menczer. 2013. Universality of scholarly impact metrics. *Journal of Informetrics* 7, 4 (2013), 924–932.

[20] Kwon Kyeongan, Majeti Dinesh, Uzzi Brian, and Ioannis Pavlidis. 2016. Scholar plot. (May 2016). Retrieved Sept. 12, 2016 from http://scholarplot.com/index.html.

[21] Bongshin Lee, Mary Czerwinski, George Robertson, and Benjamin B. Bederson. 2005. Understanding research trends in conferences using paperLens. In *CHI'05 Extended Abstracts on Human Factors in Computing Systems (CHI EA'05)*. ACM, New York, NY, 1969–1972. DOI: http://dx.doi.org/10.1145/1056808.1057069

[22] Loet Leydesdorff and Lutz Bornmann. 2011. Integrated impact indicators compared with impact factors: An alternative research design with policy implications. *Journal of the American Society for Information Science and Technology* 62, 11 (2011), 2133–2146. DOI: http://dx.doi.org/10.1002/asi.21609

[23] Evelyne Lutton, Hugo Gilbert, Waldo Cancino, Benjamin Bach, Pierre Parrend, and Pierre Collet. 2014. Gridvis: Visualisation of island-based parallel genetic algorithms. In *European Conference on the Applications of Evolutionary Computation*. Springer Berlin, Berlin, Germany, 702–713. DOI: http://dx.doi.org/10.1007/978-3-662-45523-4_57

[24] David J. Lynn, Geoffrey L. Winsor, Calvin Chan, Nicolas Richard, Matthew R. Laird, Aaron Barsky, Jennifer L. Gardy, Fiona M. Roche, Timothy H. W. Chan, Naisha Shah, and others. 2008. InnateDB: Facilitating systems-level analyses of the mammalian innate immune response. *Molecular Systems Biology* 4, 1 (2008), 218.

[25] Ian Scott MacKenzie. 2009. Citedness, uncitedness, and the murky world between. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems (CHI EA'09)*. ACM, New York, NY, 2545–2554. DOI: http://dx.doi.org/10.1145/1520340.1520360

[26] Eamonn Maguire, Javier Martin Montull, and Gilles Louppe. 2016. Visualization of publication impact. In *Proceedings of the Eurographics/IEEE VGTC Conference on Visualization: Short Papers (EuroVis'16)*. Eurographics Association, Goslar, Germany, 103–107. DOI: http://dx.doi.org/10.2312/eurovisshort.20161169

[27] Justin Matejka, Tovi Grossman, and George Fitzmaurice. 2012. Citeology: Visualizing paper genealogy. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems (CHI EA'12)*. ACM, New York, NY, 181–190. DOI: http://dx.doi.org/10.1145/2212776.2212796

[28] Teresa Penfield, Matthew J. Baker, Rosa Scoble, and Michael C. Wykes. 2014. Assessment, evaluations, and definitions of research impact: A review. *Research Evaluation* 23, 1 (2014), 21–32. DOI: http://dx.doi.org/10.1093/reseval/rvt021

[29] Adam Perer and Jimeng Sun. 2012. MatrixFlow: Temporal network visual analytics to track symptom evolution during disease progression. In *AMIA Annual Symposium Proceedings*, Vol. 2012. American Medical Informatics Association, 716–725.

[30] Filippo Radicchi, Santo Fortunato, and Claudio Castellano. 2008. Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences of the United States of America* 105, 45 (2008), 17268–17272.

[31] Lei Shi, Hanghang Tong, Jie Tang, and Chuang Lin. 2015. Vegas: Visual influence graph summarization on citation networks. *IEEE Transactions on Knowledge and Data Engineering* 27, 12 (Dec. 2015), 3417–3431. DOI: http://dx.doi.org/10.1109/TKDE.2015.2453957

[32] Ben Shneiderman. 1996. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages*. 336–343. DOI: http://dx.doi.org/10.1109/VL.1996.545307

[33] Arnab Sinha, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, Bo-June (Paul) Hsu, and Kuansan Wang. 2015. An overview of Microsoft academic service (MAS) and applications. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, New York, NY, 243–246. DOI: http://dx.doi.org/10.1145/2740908.2742839

[34] Jonathan Stallings, Eric Vance, Jiansheng Yang, Michael W. Vannier, Jimin Liang, Liaojun Pang, Liang Dai, Ivan Ye, and Ge Wang. 2013. Determining scientific impact using a collaboration index. *Proceedings of the National Academy of Sciences of the United States of AMerica* 110, 24 (2013), 9680–9685.

[35] John Stasko, Jaegul Choo, Yi Han, Mengdie Hu, Hannah Pileggi, Ramik Sadana, and Charles D. Stolper. 2013. Citevis: Exploring conference paper citation data visually. *Posters of IEEE Information Visualization* (2013).

[36] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. 2008. ArnetMiner: Extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08)*. ACM, New York, NY, 990–998. DOI: http://dx.doi.org/10.1145/1401890.1402008

[37] Robert J. W. Tijssen. 2003. Scoreboards of research excellence. *Research Evaluation* 12, 2 (2003), 91–103. DOI: http://dx.doi.org/10.3152/147154403781776690

[38] Tatiana Von Landesberger, Arjan Kuijper, Tobias Schreck, Jörn Kohlhammer, Jarke J. van Wijk, J.-D. Fekete, and Dieter W. Fellner. 2011. Visual analysis of large graphs: State-of-the-art and future research challenges. In *Computer Graphics Forum*, Vol. 30. Wiley Online Library, 1719–1749. DOI: http://dx.doi.org/10.1111/j.1467-8659.2011.01898.x

[39] Colin Ware. 2012. *Information Visualization: Perception for Design* (3rd ed.). Morgan Kaufmann, San Francisco, CA.

[40] Meng Qi Yelena Wu, Robert Faris, and Kwan-Liu Ma. 2013. Visual exploration of academic career paths. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'13)*. ACM, New York, NY, 779–786. DOI: http://dx.doi.org/10.1145/2492517.2492638

[41] Lexing Xie. 2016. Visualizing citation patterns of computer science conferences. (Aug. 2016). Retrieved Sept. 16, 2016 from http://cm.cecs.anu.edu.au/post/citation_vis.

[42] Panpan Xu, Nan Cao, Huamin Qu, and John Stasko. 2016. Interactive visual co-cluster analysis of bipartite graphs. In *2016 IEEE Pacific Visualization Symposium (PacificVis)*. 32–39. DOI: http://dx.doi.org/10.1109/PACIFICVIS.2016.7465248

[43] Wei Xu, Xin Liu, and Yihong Gong. 2003. Document clustering based on non-negative matrix factorization. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'03)*. ACM, New York, NY, 267–273. DOI: http://dx.doi.org/10.1145/860435.860485

[44] Jian Zhao, Christopher Collins, Fanny Chevalier, and Ravin Balakrishnan. 2013. Interactive exploration of implicit and explicit relations in faceted datasets. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (Dec. 2013), 2080–2089. DOI: http://dx.doi.org/10.1109/TVCG.2013.167

[45] Jian Zhao, Michael Glueck, Fanny Chevalier, Yanhong Wu, and Azam Khan. 2016. Egocentric analysis of dynamic networks with egoLines. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI'16)*. ACM, New York, NY, 5003–5014. DOI: http://dx.doi.org/10.1145/2858036.2858488

[46] Jian Zhao, Zhicheng Liu, Mira Dontcheva, Aaron Hertzmann, and Alan Wilson. 2015. MatrixWave: Visual comparison of event sequence data. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI'15)*. ACM, New York, NY, 259–268. DOI: http://dx.doi.org/10.1145/2702123.2702419

Received January 2017; revised May 2017; accepted July 2017