SuccessVis - Visualizing Academic Impact

Calvin Yau*, Chittayong Surakitbanharn†, Jun Xiang Tee‡ and David Ebert*

*VACCINE, Purdue University, West Lafayette, USA
E-mail: yauc@purdue.edu, ebertd@purdue.edu

†Civil and Environmental Engineering, Standford University, Stanford, USA
E-mail: jao@standford.edu

‡Google, Inc
E-mail: crimsonfaith91@gmail.com

Abstract—SuccessVis is a flexible web-based visual analytics system for presenting and evaluating academic impact. The system is developed for stakeholders who have little training in visualization techniques and have invested in academic programs, specifically in the VACCINE (Visual Analytics for Command, Control and Interoperability Environments) Center, to understand the impact of the center and, thus, better evaluate their future investments in academic research centers. The system is capable of displaying metrics indicating the total impact of the academic research center and allows comparisons between the impact of individual projects, and the dynamics and evolvement of the impact. In this paper, we present the design process of the system, the final product displaying the impact of the VACCINE Center over the past eight years, and discuss the challenges in visualizing academic impact.

Index Terms—Academic Impact, Communication-minded Visualization, Casual Visualization.

I. INTRODUCTION

Exploring and evaluating the success of university activities is challenging because of the various dimensions of impact (e.g., research, education, outreach). For research activities, a common approach is exploring citation networks [6]. Researchers started visualizing the connections between citations, fundings, and research topics (e.g., [1], [11], [13]) to better understand the impact of science. As the research on academic impact evolved from the traditional "scientometrics" [3] or "informetrics" [4] and became more accessible, a new group of users, "casual experts" [11], became interested. Casual experts, a term first introduced by Madhavan et al. in 2014, are domain experts in specific fields that have little training in visualization and visual analytics but are familiar with their data and want to understand the impact of their work to evaluate future directions. Casual experts often hold decision-making positions and would like to support their decisions with data and facts.

During the 7th year of the United States Department of Homeland Security (DHS) VACCINE (Visual Analytics for Command, Control and Interoperability Environments) Center, our sponsor became interested in learning about the achievements of the center and its changes over time to help better evaluate future investments in universities. However, most previous scientometric works focus on a single temporal snapshot and rarely provide insights into the dynamic temporal changes (e.g., [1], [7], [11], [13]). Moreover, the impact of academic programs can be difficult to measure compared to industry since most outcomes come in the form of discoveries of new knowledge or developments of new techniques, rather than comparable financial profit. After discussion, we decided to develop a visual analytics system to explore and evaluate the magnitude and the temporal evolution of our success, instead of compiling a traditional text-based report.

Visualizing academic impact has its challenges. Most of the data from research labs are abstract and non-numeric data such as publications, presentation slides, reports and logs from different projects. Such outputs are difficult to port into traditional visualization systems [15] that are simpler and easier for casual experts to understand. To address this problem, we quantified research impact, preserved important contextual data, and then combined simple but effective visualization techniques to create an interactive exploration and analysis system of the products and projects, their impact, and their evolvement over time. The resulting system is called SuccessVis, a flexible and easily adaptable web-based visual analytics system for examining the impact of different academic research work and how their different areas of impact change over time. SuccessVis is specifically designed for audiences who are not familiar with visual analytics to examine and compare the magnitude of the impact from different projects and different impact dimensions (e.g., research vs. real-world application) and connect these to contextual data to provide an understanding to the comprehensive high-level story. The system also provides additional supportive materials on each project and milestone for users who are interested in further exploring the successes of the center.

In the content that follows, we present the related work, our design process and the final product. We will then discuss our use case, the challenges we identified, and the feedback we collected. Finally, we will describe the potential applications and future directions.

II. Related Work

Eugene Garfield first introduced the idea of impact factor [6] in 1955 from the study on Citation Indexes for Science. Scientists and published papers within certain fields of study were ranked and compared using the impact factor. The impact factor is, however, often misused [5], and is more appropriate for comparing researchers or research works rather than understanding the impact and its dynamics. It is also limited to achievements related to publications, and academic impact should consider more than just publications.

Borner and Scharnhorst [1] reviewed different science conceptualizations used for comparing existing datasets and models. These visualizations, however, mostly focus on the relations and the connections between different topics or specific works and are not fitting for visualizing the development of impact from a specific center.

CoE-Explorer [7] was created to help DHS program managers analyze and present the research works within DHS network of Centers of Excellence, led by multiple universities, to external policy makers. The system visualizes details of individual centers, projects, and investigators. Its primary output, however, provides the general themes as an overview of the different works within the research organization, rather than a better understanding of the research works and their impact.

STAR Metrics [13] is a repository of data and tools that allows users to examine the funding and impact of federal investments across the United States. It allows users to drill down into the data by location, organization, topic, project, etc. The visualization tools serve more as filters and the data remains to be displayed in a spreadsheet and text description format. The impact unfortunately only considers the final list of publications and patents.

Similarly, Madhavan et al. [11] created DIA2 to help casual users, specifically program managers and academic staffs at the National Science Foundation, explore research funding portfolios. The system focuses on organizational structures, collaboration networks, fundings and awards received. Users can get a glimpse of the impact of a particular portfolio over the years through simple statistical graphs such as a bar graph on the number of awards received, the amount of fundings received, or the number of collaborators per year. It, however, does not incorporate contextual data and does not support analysis and comparison of the different fields of impact within a portfolio or between different portfolios.

IBM developed a free online data visualization tool, Many Eyes [12], that targets mass users with any good set of data but have no programming or technical expertise, which fits the description of the casual experts for our work. Unfortunately, the system works better with numerical data and provides the list of visualization techniques based on the data provided. Users will need to have enough knowledge on the strength and possible confusion of each technique and will be limited to only the techniques the system can provide.

III. Design Process

A. Desired Output

To present academic impact, our first step was to decide on the desired output. An interactive visual analytics system allows its users to explore large and complex datasets and obtain insights that aids decision making [14]. Being able to interactively filter and examine the center's impact in different areas visually can be more effective than examining numbers and text over spreadsheets or paragraphs when exploring the outcome of investments. Our goal is to, through providing a visual analytics system, effectively communicate the scale, the temporal changes, and the context of the impact from the VACCINE Center in different areas and provide an intuitive visual analytics experience for casual experts to obtain insights and support decision makings.

B. Data Compilation

Determining what data was available and categorizing it was a key step in our design process. Most of the data we were able to collect from our center universities and research came from our annual reports to the sponsor. Through the reports, we obtained information such as the publications produced, important presentations, patents received for each project and lists of funded students from each partnering universities. There were also monthly newsletters and lists of seminars and presentations from the center, as well as different presentation slides, flyers, video demonstrations, software systems and logs from the different projects. These were mostly non-numeric contextual data.

To visualize this available contextual data using more commonly understandable techniques that typically requires numeric data, we needed to separate the data into quantifiable data, contextual data, and supportive materials. Quantifiable data are variables whose significance and magnitude are positively correlated and can be represented in or transformed into a countable manner. For example, the amount of funding a center obtains may be proportional to the outside interest in its research work. Quantifiable data can be directly supported by traditional visualization techniques that are common to general audiences, including line graphs and bar charts. Contextual data are variables whose significance are not directly represented by their count and require more content to provide insight. For example, a prestigious award might be more significant than multiple less prestigious awards. Contextual data can be presented using short text descriptions, annotations, or glyphs – representations that do not require extensive training in visual analytics to understand or use. Finally, supportive materials, such as a recorded demonstration of a working system or the paper from a research discovery, provide additional details into the story for the impact variables.

From the list of publications, we acquired the number of publications per year for each project, and then determined the number of citations per year. From VACCINE's annual reports and quarterly newsletters, we estimated the number of important presentations. From the system logs, we either obtained

the exact number of users based on the accounts created or estimated the number of users based on the unique download IP addresses. Given the variety of sources and media and activity knowledge needed, human intervention was needed to generate the statistical data. Tools such as Wrangler [9] may be able to reduce the degree of human-effort and make this process more accessible. We also collected contextual data such as project milestones, start dates and end dates, significant project transitions, workshops hosted, competitions participated and awards received from annual reports and newsletters. The contextual information would be presented using short text descriptions and glyphs and connect to the visual components by date and project. Finally, there were additional supportive materials such as posters, presentation slides, and videos that could be attached to relevant textual data for further investigations.

Since we aimed to present multiple projects from the center, we needed to find the data fields commonly shared, and therefore comparable, that could represent the scale of impact. After examining the amount of data from each data field for our main projects, we finalized on four fields or areas of impact: publication counts, citation counts, presentation counts, and user counts. All four fields, besides user counts on rare occasions, are fairly common across the different types of academic projects and cover a variety of impact types. The number of publications reflects the novelty and depth of the research work, the number of citations reflects the relevance to other research works, the number of presentations reflects the interest from the outside world, and the number of users reflects the practicality of the research output.

C. Visualization

With the collected data grouped and organized, our next step was to design a visual display that can incorporate and present the three groups of data (quantifiable, contextual, and supportive) we collected for the different projects. We focus our visualization mainly on the quantifiable data, encoding the magnitude and time span of the different impact types. Since we are not only interested in the final comparison, but also on how the different types of impact change over time, we considered multiple time-series-focused techniques to visualize our data.

We selected a stacked graph using the ThemeRiver layout [2] to visualize the scale of impact and its changes over time, as the strength of ThemeRiver graph lies in visualizing thematic variations in both individual topics and groups of topics over time [8]. Such strength fits our need in showing both the impact of the center as a whole, the impact of individual projects, and the different types of impact from each project in a measurable and comparable manner. Being able to see the impact in different areas for different project categories is helpful for investors to evaluate the direction and the timespan of their next investment. ThemeRiver is also similar enough to traditional line graphs for casual experts to use effectively.

We then combined the idea of project timeline into ThemeRiver to incorporate the milestone information that is connected to the quantifiable impact by time and project. To further support the visualization with contextual data we created an interactive system that presents the data and provides additional information on the projects and more details and supportive materials on the milestones. A hover-highlight functionality is added to connect the substreams to the high-level project story and address the challenge in scalability as the number of substreams increases. Glyphs and tooltips with short text description are also overlaid on top of the streams to connect to the detailed stories.

D. User-Centered Design

We designed our system with an iterative and incremental development approach [10]. We first presented a mock-up to our sponsor. From their feedback, we created a prototype system. After the second iteration of feedback with the prototype, we refined the design and populated it with three major projects from the VACCINE Center. With the following feedback, we populated the full data and continued demonstrating our refined system to our sponsor then incorporating their feedback back into the system. To ensure the design appeals to casual experts, we also reached out to a Purdue University Interaction Design & Industrial Design team to improve our color choices and glyph designs for the milestone icons used in the system.

We distributed the work on collecting and filtering the milestone contextual data to administrative staff, instead of the researchers, to minimize the possibility of bias on what would be considered as important stories to tell. For example, researchers see the discovery and overcoming of roadblocks during a project as important achievements, while sponsors pay more attention to project launches, transitions, and awards. It is important that the information our system provides is relevant to the audience rather than the presenter.

Finally, to make the system usable over time as center activities continue, and to make it beneficial to other centers without redevelopment, we decided to create a system that is simple to update and flexible to adapt to different fields and measure. We also decided to make the system web-based to avoid installation issues and enable easy access from anywhere for use and presentation. The system stores data and populates the user interface through spreadsheets for ease in updating, exporting from databases, etc.

IV. SYSTEM

Figure 1 displays the system consisting of three parts: (1) a Project Slideshow (Figure 1a) that gives summary highlights for each projects, (2) an Impact Stream Graph (Figure 1b) that displays the magnitude of the impact and its changes over time for the different projects or types of impact, and (3) a Milestone Details View (Figure 1c) that allows users to read the milestone story in more detail and access the attached supportive materials.
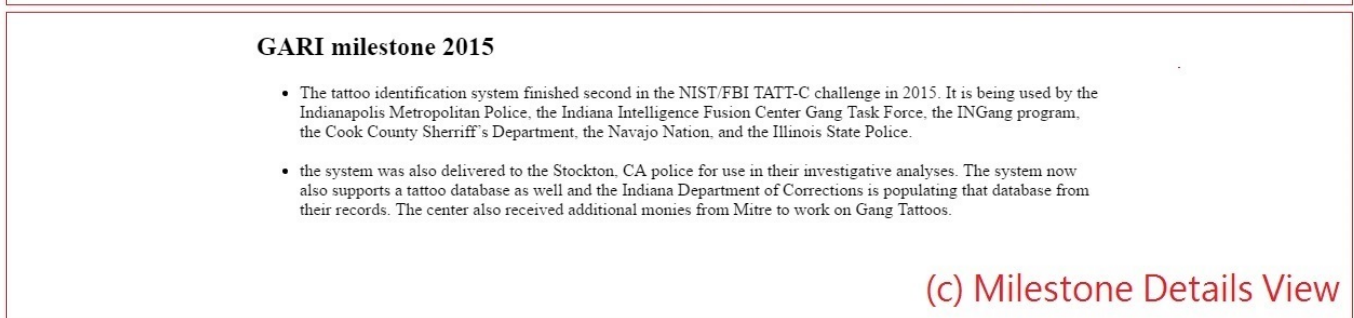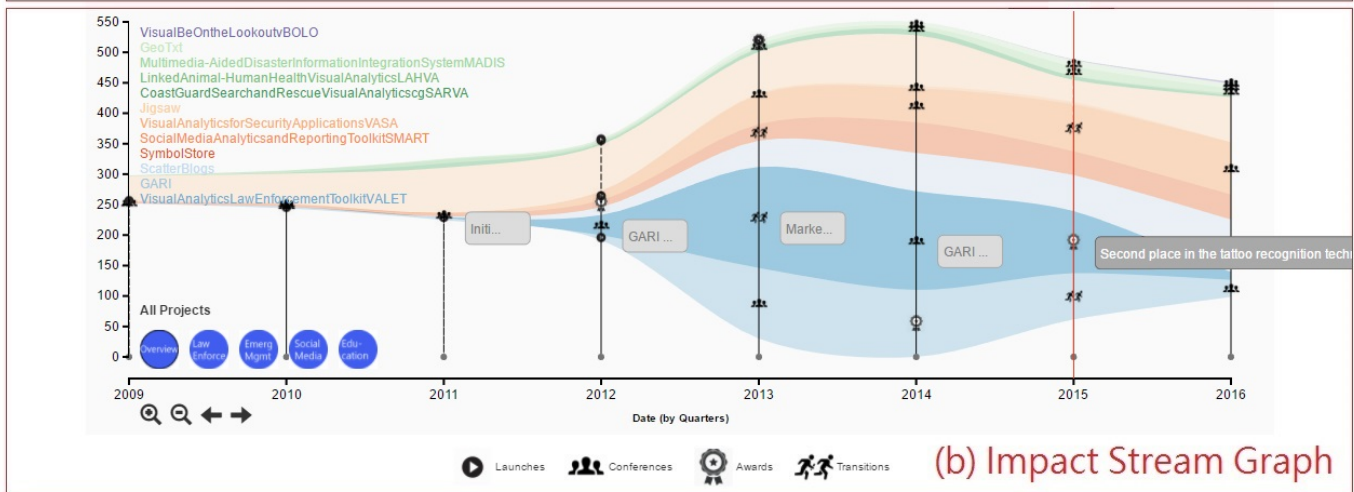
Fig. 1. The SuccessVis system contains three connected views.

A. Data Spreadsheet

To populate the system, a few specific spreadsheets have to be generated:

- slideshow.csv stores the Project Slideshow information. It includes, for each project, its project name, start date, end date, project description, and an additional media link.
- gauges.csv stores the subjects and the values of the three takeaway values on the right side of the Project Slideshow.
- metrics.csv stores an index number and a corresponding name for each tab the Impact Stream Graph displays.
- visualization_data_x.csv, where x is a positive integer, stores the actual data of a tab in the Impact Stream Graph.

Its fields include a topic key, a topic value, a date, a milestone summary, a milestone glyph, and a milestone link.

- milestone_legend.csv stores the paths to the milestone glyph image files and their corresponding milestone categories, then uses the data to generate the legend at the bottom of the Impact Stream Graph.

B. Visual Analytics

The three components of SuccessVis (Project Slideshow View, Impact Stream Graph, and Milestone Details View) combine and link the numeric data, the contextual data, and
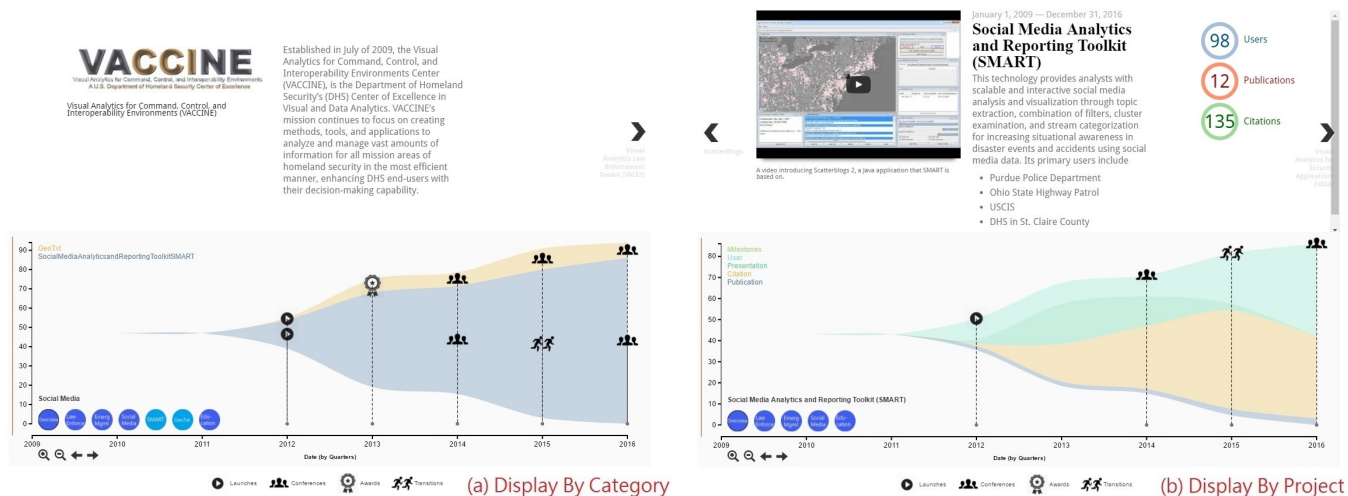
Fig. 2. SuccessVis displaying the impact breakdown of (a) the Social Media Category and (b) the project SMART.

the supportive materials to provide a full picture of the center's impact:

1) Project Slideshow View: Project Slideshow View includes three subcomponents: project description, external media, and takeaway values. The project description, displayed in the middle of the Project Slideshow, provides a summary of the purpose, the output, and the partners of the project. The external media, displayed on the left side of the Project Slideshow, is capable of displaying an image, a video, or a PDF file as long as a valid link address is stored in the spreadsheet. The takeaway values are located on the right side of the Project Slideshow. It provides the users with values on the project as a whole. The slideshow starts off with an initial slide on the center itself, then lists the projects ordered from left to right in chronological order. Users can click on the left or the right arrows to navigate through the projects. A fainted summary that includes the project name and timespan of the next/previous project, located under the arrows, will be highlighted when users hover the mouse over the arrows. The project the slideshow is displaying will also be highlighted on the Impact Stream Graph if it is included in the selected tab.

2) Impact Stream Graph: Impact Stream Graph displays the impact of the center's work. It is capable of displaying multiple tabs of impact, which are selected through the blue circle icons on the lower left corner, allowing users to display and examine impacts from different projects and categories. Tab one displays all the projects' collective impact, as seen in Figure 1b, showing how the impact from the center as a whole has evolved. The substreams represent the total impact of each project showing the proportion of their contributions to the overall impact of the center. The x-axis encodes time, while the unit for the y-axis depends on the subjects or metrics displayed. The two axes enable quantitative comparative analysis. While examining the VACCINE projects, we grouped the projects into different categories such as law enforcement, resource allocation, and social media. In

Figure 2a an user has selected the Social Media tab, leading the Impact Stream Graph to display the collective impact of GeoTxt and SMART from the Social Media category and the proportion they each contributed to the combined magnitude. By double clicking on the larger substream that represents SMART or by clicking the SMART tab icon, the graph will display a project impact breakdown with the substreams being the different impact fields, as shown in Figure 2b, allowing the user to better understand what areas the projects have a stronger impact on. While the system currently does not support adding weights directly to the different metrics or projects, weighted comparison or direct comparison between specific metrics can be easily added as new tabs with new spreadsheets.

Different icons are placed on top of the stream graph at the corresponding time to show important milestones each project achieved. The different glyphs represent different categories of milestones such as project launching, transitions, and awards. By hovering over a milestone icon, a summary of the milestone story will be displayed. By clicking on the milestone icon, a detailed description of the milestone will be displayed in the Milestone Details View, as shown in Figure 1c. When a project is highlighted, either through the slideshow or through hovering the mouse over the substream, the milestone icons belonging to the project will be highlighted to help users better connect the projects and the milestones.

3) Milestone Details: Milestone Details View provides a more in-depth description of milestones. It also displays other important events that happened during the same timeframe that are unable to be displayed on the the milestone summary in the Impact Stream Graph. Milestone Details View takes a web URL link and displays it in an iframe when its corresponding milestone icon is clicked. By displaying a webpage, it is capable of not only explaining the milestone stories in more detail, but also embedding images, videos, external links, or other attachment files for users to examine.

V. Discussion

A. Use Case

We used SuccessVis to visualize the impact of VACCINE Center from the past eight years. We selected 11 major projects to display in the system to learn about the strengths and weaknesses of the different projects and project categories regarding impact. For example, by examining the law enforcement category, we can see three of the four projects have more users compared to most of the projects, indicating that the law enforcement projects have a strong impact in the work field. We can also examine how the proportions of different impact fields from a project changed as time progress. For example, In Figure 2b, we can see how around the year 2015, SMART's users started increasing more dramatically while the number in presentations and citations started decreasing, indicating the overall impact transitioned from a more theoretical interest into more practical uses after about three years of development. By performing similar analyses across different projects among different categories, our sponsor can better evaluate and consider the progress in the field and timespan for their future investments. Moreover, they can identify successful project patterns for evaluating new projects that may appear to be progressing slowly.

B. Generalizability

Our system has been designed to be adaptable to other academic centers. Three out of the four impact types are collected from common products of most academic research centers. The user count and other possible impact factors not included can be easily added or removed using our populate-through-spreadsheet method. While the system does not currently support automated source material conversion directly, the process can be replicated manually and is easily semi-automated through periodic data collection, rather than gathered at the end of the period of interest.

C. Challenges

In spite of our initial system, there remain challenges to be solved such as:

- Data frequency: Paper publications and citations often happen at the rate of one every few months, while new users and presentations vary between the different projects. If we input the stream graph data by month, we could end up with multiple zeros, which creates unattractive visuals and may not present the impact of the publications and citations fairly. However, when we input the data by year, we can only display one milestone icon per project per year due to the limitation of the spreadsheet-filled system. Multiple milestones of different kinds could occur within the same year, and it is difficult to represent the different events with just one icon.
- Limited usable data fields: To ensure reasonable comparison between the impact magnitudes of different projects, the project impacts have to share the same quantifiable data fields. We thus are limited to only the data that exists for all the projects we plan to display.
- Combining impacts: To compare the impacts from different projects, we need a representative overall impact value for each project. To not be biased toward a specific field, we add up the four different fields of impact to get an estimated project impact. However, the four fields do not share the same units or standardized weights, and adding up the fields makes the actual measurement confusing. If we instead normalize the different fields before adding them up, the comparison between the project magnitudes will not be accurate.
- Project duration: Not all the projects share a similar amount of data. A simple spreadsheet may be more effective for short projects than the system. But how to decide on the threshold is difficult.

D. Initial Feedback

After initial presentations of our system with VACCINE's eight-year data, our sponsor, the Science and Technology directorate of the United States Department of Homeland Security, indicated appreciation for the storytelling aspects of the system and the ease to understand and compare the overall picture of the center and its projects. The sponsor also expressed interest in having us provide a system template and user guide to distribute to other sponsored academic research centers to showcase their center impact. The sponsor was interests in learning more about the educational impacts. To respond to this request, we collected data regarding courses and students benefited from the fundings and inserted it into the system as an additional and standalone project.

We had the chance to present this work to faculties from other universities. Many of them discussed with us the possibility to access the system and the opportunity to populate their own research impact. We hope to collaborate with them in the future to further evaluate and improve our system.

One feedback expressed concern about a potential color matching issue between the takeaway values in the Project Slideshow and the impact type in the Impact Stream Graph. This is, unfortunately, a result of the approach to populating the system through dynamic spreadsheets. Linking the repeated fields between the takeaway values in the Project Sideshow and the fields in the Impact Stream Graph is currently not supported. We aim to resolve this issue in future work.

VI. Conclusions and Future Work

In this paper, we presented a visual analytics system for evaluating the magnitude and temporal evolution of academic impact. The primary target audience of this system includes stakeholders who want to evaluate their future investments in academic research. We grouped the unconventional and unorganized data collection from a research center into quantifiable data, contextual data, and supportive materials and visualized the three groups of data using a Project Slideshow, Impact Stream Graph, and a Milestone Details View. The result is

a visual analytics system that summarizes the impact stories of an academic research center in a manner that allows high level comparative analysis between different subjects while preserving the context and providing more detailed stories on demand for stakeholders who have little training in visual analytics.

We plan on continually refining our system based on the feedback from our sponsor. We also plan on continuously collaborating with the Interaction Design & Industrial Design team to add a project tree view that lists all the projects from the center instead of limiting to major projects. Such tree will present the timespan and parent-child relationships of each project. By hovering over a branch on the tree, a tooltip will appear to provide the basic information of the corresponding project. The tree will display a complete picture of all the projects from the center, including projects that do not have enough data for the current system and provide the additional parent-child information of different projects that the current system is not capable of presenting. Future research should explore methods for combining impact fields of different areas and presenting impact fields of different frequencies. In addition to the technical aspects, we also plan on providing the system to other research centers and evaluate how well the system adapts to their data and how useful the system is to researchers without programming or visual analytics backgrounds.

REFERENCES

[1] K. Brner and A. Scharnhorst. Visual conceptualizations and models of science. *Journal of Informetrics*, 3(3):161 – 172, 2009. Science of Science: Conceptualizations and Models of Science. doi: 10.1016/j.joi.2009.03.008

[2] L. Byron and M. Wattenberg. Stacked graphs–geometry & aesthetics. *IEEE transactions on visualization and computer graphics*, 14(6), 2008.

[3] D. de Solla Price. Editorial statements. *Scientometrics*, 1(1):3–8, 1978.

[4] L. Egghe and R. Rousseau. *Introduction to informetrics: Quantitative methods in library, documentation and information science*. Elsevier Science Publishers, 1990.

[5] E. Garfield. The impact factor and using it correctly. *Der Unfallchirurg*, 48(2):413, 1998.

[6] E. Garfield. Citation indexes for science. a new dimension in documentation through association of ideas. *International Journal of Epidemiology*, 35(5):1123, 2006. doi: 10.1093/ije/dyl189

[7] S. Ghani, N. Elmqvist, and D. S. Ebert. Multinode-explorer: A visual analytics framework for generatingweb-based multimodal graph visualizations. *Proceedings of EuroVA*, pp. 67–71, 2012.

[8] S. Havre, B. Hetzler, and L. Nowell. Themeriver: Visualizing theme changes over time. In *IEEE Symposium on Information Visualization*, pp. 115–123. IEEE, 2000.

[9] S. Havre, A. Paepcke, J. Hellerstein, Joseph, J. Heer. Wrangler: Interactive visual specification of data transformation scripts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3363–3372. ACM, 2011.

[10] C. Larman and V. R. Basili. Iterative and incremental developments. a brief history. *Computer*, 36(6):47–56, 2003.

[11] K. Madhavan, N. Elmqvist, M. Vorvoreanu, X. Chen, Y. Wong, H. Xian, Z. Dong, and A. Johri. DIA2: Web-based cyberinfrastructure for visual analysis of funding portfolios. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1823–1832, Dec 2014. doi: 10.1109/TVCG .2014.2346747

[12] Many eyes. http://www.boostlabs.com/ibms-many-eyes-online-data-visualization-tool/. Accessed: 2017-06-14.

[13] Star metrics. https://www.starmetrics.nih.gov/. Accessed: 2017-06-14.

[14] J. J. Thomas and K. A. Cook. A visual analytics agenda. *IEEE computer graphics and applications*, 26(1):10–13, 2006.

[15] C. Upson, T. Faulhaber, D. Kamins, D. Laidlaw, D. Schlegel, J. Vroom, R. Gurwitz, and A. Van Dam. The application visualization system: A computational environment for scientific visualization. *IEEE Computer Graphics and Applications*, 9(4):30–42, 1989.