DESIGNING DATA VISUALIZATIONS FOR OPEN SCIENCE

Chun Wang
University of Virginia
21921 Providence Forge Drive,
Ashburn, VA
cw8df@virginia.edu

ABSTRACT

Web-based Open Science tools[1] provide rich sources of information about a scholar's research dynamics and increase visibility to collaboration opportunities on research projects. This paper describes the principles for designing a user interface for Open Science software that facilitates knowledge discovery in open science community and assists a visual understanding of scholars' research contributions. Our understanding of these issues is shaped by our experiences designing the user profile pages of the Open Science Framework[2], which can also serve as a case study in the practice of extracting and presenting meaning from faceted, interlinked data.

Author Keywords

Open Science, user interface, data visualization, design guidelines

1. INTRODUCTION

In an attempt to make science more "open", an increasing number of scientific researchers are turning to web-based tools to publish their research papers and useful information pertinent to the research but traditionally not journal-worthy, such as research workflows, datasets, code bases, and preliminary observations. Expanded scope of accessible data opens opportunities for both critique and collaboration in science, and facilitates would-be research contributions. However, as the volume and variety of published data grow, users of Open Science tools face unique challenges to make sense of the data, including:

- *How to easily manage my projects on the website?*
- *How to cope with the flood of published research to find materials pertinent to my research?*
- *How to stay relevant with the hot topics of my discipline and select important research problems to attend to?*
- *How to allow other scientists to view a research profile of me, including my work history and research interests?*
- *Is it possible to incorporate a measure of scientific productivity of a scholar, based on the quantity and quality of his published research data?*

The solutions to these issues primarily fall on the user interface layer of Open Science software. Software that can resolve these challenges by providing a effective user

[1] Examples of Open Science tools becoming mainstream in the community of scientists include: ResearchGate (www.researchgate.net), academia.edu (http://www.academia.edu), figshare (http://figshare.com), and Mendeley (http://www.mendeley.com). A list of emerging tools for open science can be found at http://science.okfn.org/tools-for-open-science.

[2] http://cos.io, developed by Center for Open Science, Charlottesville, Virginia.

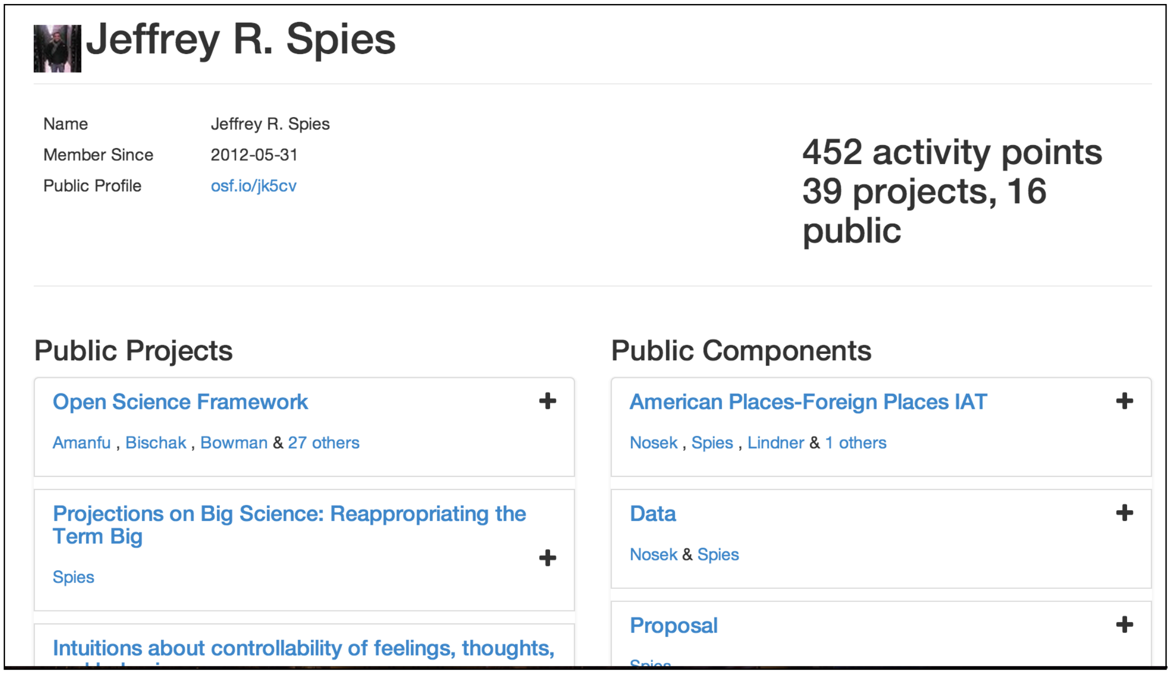

Figure 1. A sample user profile from OSF at start of project.

interface will allow scientists to leverage the benefits of Open Science, and provide the motivation to adopt Open Science tool as an integral part of their research routines. Therefore, in our design process, we have made addressing these issues as our primary design goals. To explore how a visual interface could support the desired functionality, we carried out a design process through multiple iterations. In this paper, we provide a mix of principles and practical advice, illustrated by the process behind our attempt to re-design the user profile page of the Open Science Framework (OSF). We first describe the complexities of data and which subset of data we worked with. Then we talk about our initial prototypes which gave us insights to some important design principles for next design iterations. In part 3 we present and discuss in detail about our final design. In the process we highlight design issues we experienced as well as successful design decisions.

2. DESIGN PROCESS

The information space of Open Science Framework has a rich structure. Essentially, user activities on the OSF create a complex network that can be thought of as having two types of nodes, one being users, another being open science projects. A link can drawn between a user node and a project node if the user is a contributor to the project. Yet connections may also exist between two user nodes because they collaborate on one or more projects, or between two project nodes because they may have common contributors, shared keywords, or a forking or referencing[3] relationship. The underlying inter-connections provide many useful insights to

[3] On OSF, a project fork happens when users take a copy of source files from an existing project and build a new project upon the copied files. Users can also link/reference to a project from another project without creating a copy.

research networks, potential collaboration opportunities, and the impact of a user or a project in the community. Beyond that, a range of additional attributes about the nodes themselves are stored in backend database. For example, a typical project has dates, files, statistics, activities etc. A user has activity point[4] and activity logs. These types of information allow us to learn about the development history of projects and research activeness of scholars.

To extract and present meaning from the interlinked, faceted data, we need to first specify what subset of the data we want to display to users. As mentioned above, a user profile page serves two ends: 1) providing a portal for user to view his projects; 2) allowing profile viewers to see descriptive details about the user's research interests, contributions, and the impact of his projects. At the beginning of project, OSF had a plain user profile page consisting mainly of a list of projects to which the user contributed (Figure 1). We wanted to reserve this main functionality, but also populate the page with more pertinent information that served our second purpose. We identified the key areas to include in visualization:

User-based:

- Proportion of public projects: a measure of "openness" of user activity
- Contributed projects: allow viewers get an overview of the user's contributions, and for project management purpose
- Collaborators: facilitate networking and collaboration

Project-based:

- Creation date/update dates, activity log: to get a user work history
- Category: useful information about the nature of projects
- Contributors: relate users to project and for networking purpose
- Total number of contribution: a measure of project significance and activeness
- Keywords: help users locates pertinent research materials more easily

DESIGN PRINCIPLES

We then moved on to working for the presentational aspects: how to visually display the data effectively, and what user interactions should be involved. In this section we discuss a set of design principles that we learned from the observations of our initial prototypes and selected examples from related work.

Cohesion

Our initial design involved creating a full page of distributed visualizations with each subgraph working with a small subset of data features, to hopefully present some interesting story about the user and about the community. Related work such as GitHub[5] project page [*Appendix* Figure 1] and ResearchGate[6] user profile page [*Appendix* Figure 2] have utilized this approach. Our attempts included creating themed graphs that encode data features in diverse visual elements. They are generally project-based (Figure 2 (a)), collaborator-

[4] a measure of user activeness on OSF. A point is awarded when user performs a pre-defined set of operations such as creating a project, making projects publicly viewable, uploading files to projects, etc.

[5] www.github.com

[6] www.researchgate.net

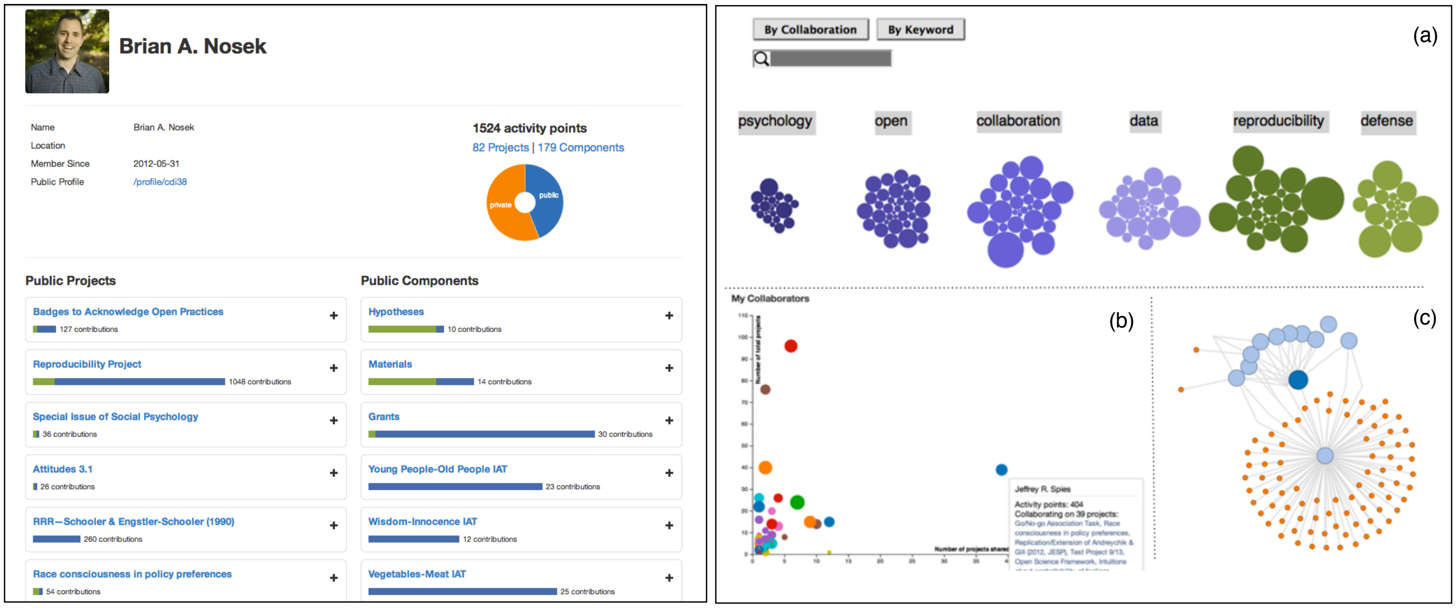

Figure 2. Initial design for OSF user profile page, featuring a project list and a separate section for themed visualization. Left: Visuals attempt to depict user's "open" contribution level. The donut shows how much percentage of user's projects are public and bars measure user contributions v.s. total contributions to a project. **Right:** A set of prototypes focused on node-based representations. (a) a keyword cluster for the user, with each node representing a project (b) collaborator nodes in a plane. x-axis: number of projects the user shares with a collaborator node, y-axis: total number of projects that collaborator contributes to (c) a network view of the user's projects and collaborators. In each of the graphs, node sizes are determined by user activity points or project

based (Figure 2 (b)), or network-based (Figure 2 (c)). The intention was to reveal patterns about some key factors regarding a user's research activities such as his networks and research interests. We believed that viewers could utilize web browser as the context for exploring information spaces and read the graphs with a casual mindset [1]. Although themed visualizations effectively suppressed noise, they could not efficiently form a coherent whole due to their distinct natures. Learning to interpret each of them appeared as a inundating task to viewers. Further, these visualizations were not suitable for serving the end of project management because they each showed only a subset of all projects. What users preferred was a portal for displaying all their projects in a coherent and consistent way.

Interpretability is related to a right amount and type of visualization

Unsuitable or unfamiliar abstractions may impede interpretation. Ideally, model abstractions should correspond to analysts' mental models of a domain to aid reasoning [2]. Because most information we consume comes in verbally rather than visually, increasing the amount of uncommon visualization will highlight the chances that viewers start to feel unfamiliar with and incompetent to interpret the presentation. The prototypes we initially created were fully visual-centered and utilized relatively uncommon elements such as clusters and networks. This practice increased the level of difficulty for interpretation. In the next design iteration, we considered incorporating more conventional visual elements such as search box, filter box, and so on. We also attempted to reduce the "visual dimensions" of an element we used (for example, nodes may not vary in size to alleviate visual burden).

Efficiency

Efficiency refers to both the time within which users find the information they need

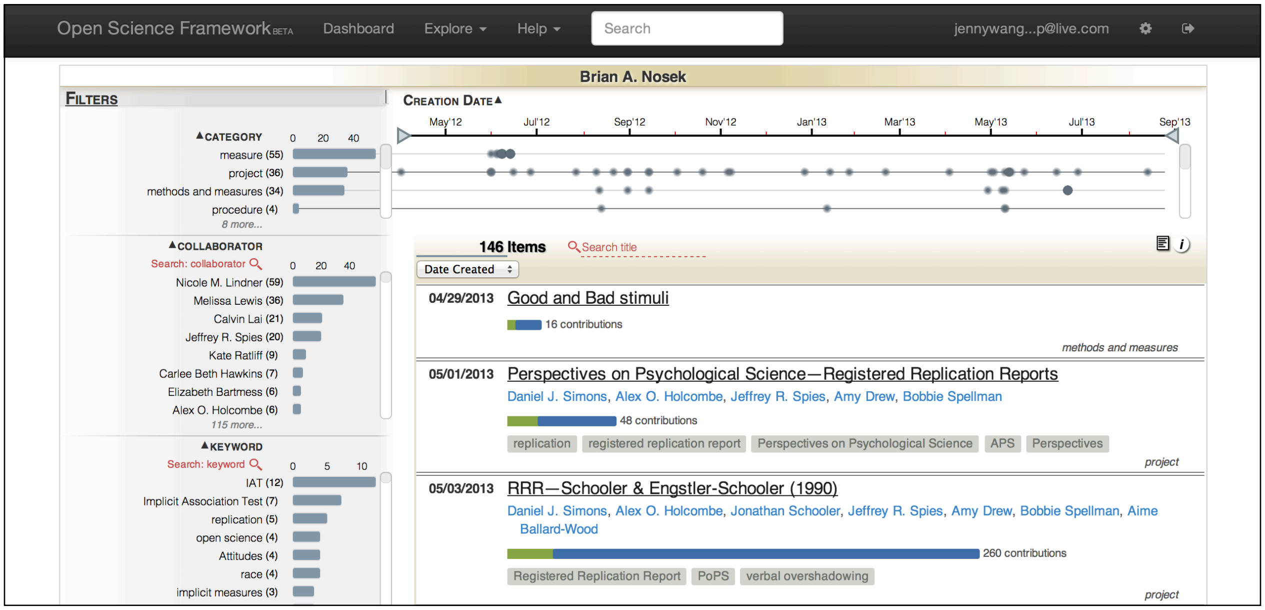

Figure 3. Final interface design for OSF user profile pages.

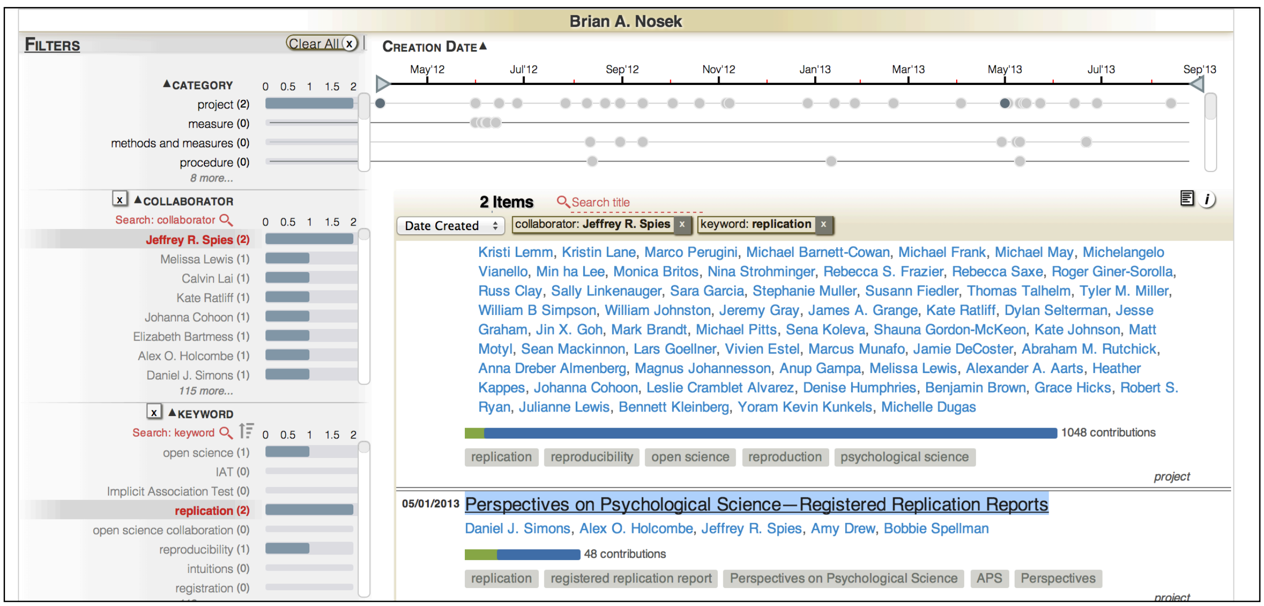

Figure 4. Main display area changes its content based on filter values. In this case, we wanted projects with “Jeffrey R. Spies” as a collaborator and “replication” in keywords.

and space efficiency. An intuitive organization of information and simple user interaction design are directly related to time efficiency. An often ignored problem of visualization is space efficiency. Standalone graphs typically take up a large portion of screen space, while the information they encode can often be expressed in a much more space-efficient way using words and figures. For this purpose we decided to eliminate using standalone graphics and instead utilize a combination of verbal phrases, figures, and visuals to present data, where verbal and logical representations had dominant place. In this way, the overall interface also achieved a desired “minimalist design” [3].

Interactivity

In the era of Web 2.0, interactivity of web applications is a familiar concept and interactivity, users can enjoy the freedom and sense of control when navigating the interface and exploring the information space themselves. Web interactivity is an appealing design element, but it is also functionally important for open science tools. The main interactive capability is to help users cope with large quantity of published research materials by enabling filtering and searching under user-defined criteria, and allowing users to navigate to the linked resources.

3. IMPLEMENTATION

In this section, we present our final design for OSF user profile and discuss its implementation aspects.

FINAL PRODUCT

The main product of our final design is an interactive project browser equipped with filtering, searching, sorting, scrolling functions and minimalist visual aids [Figure 3]. We took the inspiration from an open source project, Keshif[7], and built up more functionalities based on their design to flesh out our final product. Here is a list of main functionality that our browser provides with the goals of each element explained.

Timeline

A timeline of all projects that user contributes to (represented by nodes) is provided as a visual cue for the user's work history and contribution frequency.

Filters

A group of filters are located on the left side to show a ranked list of user's project types, collaborators, and project keywords. The intention is to give profile viewers a clear picture of what nature of work the user does, who most frequently collaborate with the user, and what his research interests are. Search bars are provided to facilitate quick lookup in. Just as clicking on timeline nodes will trigger the main display area to show projects created within a certain period of time, clicking on any specific element in the filters will result in main box displaying projects that satisfy selected criteria. These filters essentially present the same information that we wished to achieve with network graphs - for example, we wanted to visualize the most frequent collaborators - but are far more space-efficient and maybe more aligned with users' mental model of logical reasoning.

Main display area

A list of projects with their details are displayed in the On page load, this area contains all of the user's projects, but its content changes with any filters applied (see Figure 4 for example). The projects are listed in a sorted order by date created, but other orders are specifiable as well. Queries are shown at the top of box, which can be cleared or reconstructed. A minimal amount of visualization is incorporated to each project item to help viewers identify user contributions (green bar width) v.s. total number of contributions made to that project (blue bar width plus green bar width). Project titles and collaborator names link to their specific home pages to facilitate further exploration. An omni search bar is also provided to give user more freedom of exploration.

PERFORMANCE & EVALUATION

All of the browser's behaviors, including data aggregation functions for the filters, are defined in a JavaScript file. The OSF uses a MongoDB instance for data store, and to comply with current infrastructure of OSF,

[7] GitHub repository: github.com/adilyalcin/Keshif, authored by Ph.D. student M. Adil Yalçın.

we define a URL route for holding all relevant data needed for the visualization in JSON format. On each profile page load, the JavaScript will dynamically parse the JSON from the designated URL and create a browser instance accordingly. To reduce render-time, we have pre-stored additional project data in our database. For example, we have pre-calculated the widths of contribution meter bars based on our knowledge about user contributions and stored them in JSON designated URLs. The overall performance of our visualization is satisfying, with an average of 1.2s of response time on browser load and responding to filter queries almost instantly. Although not all of our design's eventual uses can be presently foreseen, the architecture provides good extensibility for future addition of new data, features, and functionality.

Is the interface well suited for the tasks? Each function of the interface contributes to the design goals. Our final design makes a one-stop portal for both project management and profile presentation. Compared with the initial user profile page, our design allows users to navigate through projects much more easily. Compared with our initial graphic-centered prototypes, the final design is more space efficient and has a better response-time performance (since there is minimal amount of graphic generation), while providing the same level of details in information presenting and higher interpretability. Its ambitious usage of filters greatly facilitates location of pertinent materials from a swarm of information. The contribution meters give viewers an overview of project significance while provide detailed information about user's research efforts. Coupled with an interactive timeline, they allow viewers to see a work history of the user.

4. CONCLUSIONS

To assist the understanding of meanings behind the large quantity of data collected by OSF about open science projects and scholars, we developed an interactive interface to help users organize their research materials and make sense of the available data. We offered a series of design considerations for constructing an easy-to-use tool that meets a series of functional goals and design challenges. Our design allows us to take user data one step further, and tease out interesting stories about users contributions and projects in open science through a presentation layer of the application. Future work on improving this interface design should include more extensive usability tests and user evaluations. There exist future design possibilities, such as visualizing the forking/referencing relationship between projects, and creating more filtering/sorting options. We encourage organizations that are developing open science platforms not only provide efficient tools, but also think in terms of creating incentives and rewards for users of their tools. We also believe that open science software development teams should have the wider scientific community join their software design process to identify the incentive and technological problems they face in publishing, discovering, and reusing research online, so tools can be improved to meet their needs.

REFERENCES

[1] Dork, M., Nathalie Henry Riche, G. Ramos, and S. Dumais. "PivotPaths: Strolling through Faceted Information Spaces." IEEE Transactions on Visualization and Computer Graphics18.12 (2012): 2709-2718. Print.

[2] Interpretation and Trust: Designing Model-Driven Visualizations for Text

Analysis Jason Chuang, Daniel Ramage, Christopher D. Manning, Jeffrey Heer ACM Human Factors in Computing Systems (CHI), 2012.

[3] NIELSEN , JAKOB . "Nielsen Norman Group." 10 Heuristics for User Interface Design: Article by Jakob Nielsen. Nielsen Norman Group, 1 Jan. 1995. Web. 25 Apr. 2014. <http://www.nngroup.com/articles/ten-usability-heuristics/>.

APPENDIX

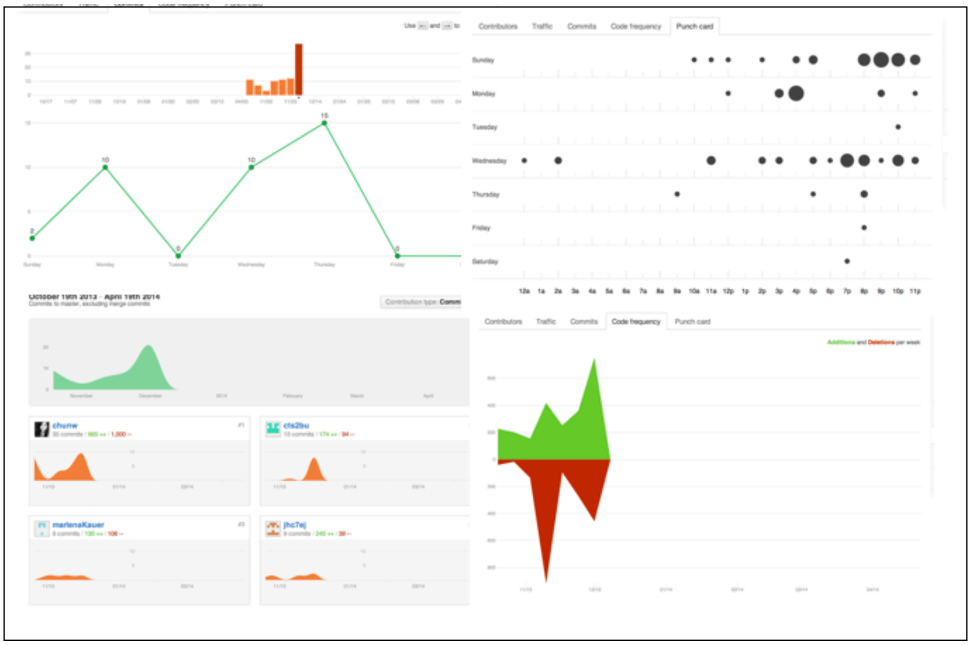

Figure 1. GitHub project graphs depict contributions from different perspectives. The graphs are visually separated under different tabs such as "Contributors", "Commits", "Punch Card", etc.

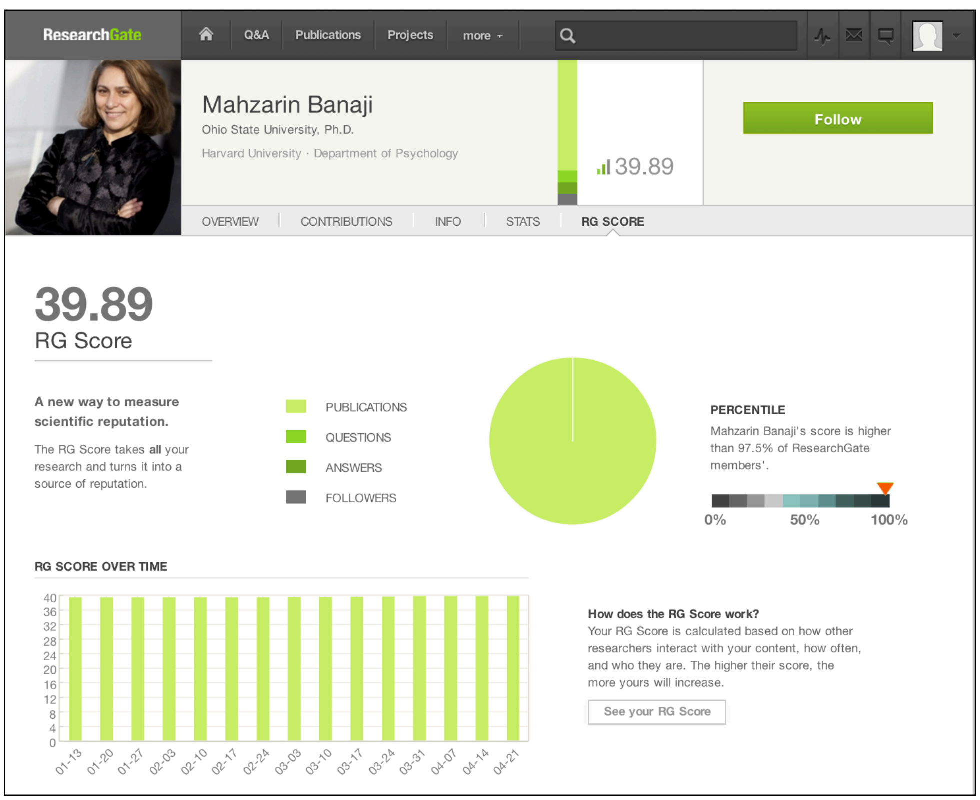

Figure 2. ResearchGate user profile visuals for RG Score related information.