

Image conscious: Visualizing scholarly reputation

William Bejarano

Judit H. Ward

*Center of Alcohol Studies,
Rutgers, The State University of New Jersey*

In an increasingly data-driven world, the amount of information available on any given topic can begin to feel overwhelming, especially to those of us who work in the information profession. This paper gives an overview of how displaying information in a visual way can turn large, complex data sets into something more digestible and understandable to the general public. The foundational principles of human visual perception are established, which should be applied to any attempt to visualize data, and a brief tutorial on the process and output of select tools is offered. To show its practical relevance, examples are presented of how the concept of “scholarly reputation” is uniquely visualized in various platforms.

Keywords

Visualization, Academic achievement, Scholarly social media

There are essentially three ways to accurately represent and communicate an abstract concept. The first is with the use of language: depicting concepts as words, strung together to form sentences, which can be placed in a specific order to be published in, say, a conference proceedings. The second, and more universal, method is the use of data tables. When using this method, a concept can be correlated with another concept and placed on an organized chart. Often, the concepts are quantified in some way, and with that, numbers can reach across languages and thus extend beyond cultural hang-ups which are intrinsically tied to linguistic expressions of concepts. The third method, and the focus of this paper, is taking data one step further with the use of

visualizations, which tap into the realms of the human perceptual system to pre-cognitively disseminate an idea.

Statistician and visualization pioneer Edward Tufte conceived some very basic principles to keep in mind when analyzing or creating visual displays. The first is to keep the focus on the data. In his language, all “ink” (pixels in an electronic display) should be used to display the data, and he considers superfluous images antithetical to this principle. Second, the simple image should be able to show as much data as possible without overwhelming the user, an idea that will be explored in more detail later in this paper. Third, the “lie factor,” when proportions and sizes of images do not correspond with the data they are

representing, should be minimized. For example, if one data dimension is twice as large as another, the resulting image should be twice the size (Tufte, 1983).



Figure 1. Bad visualization (source: [SmashingMagazine](#)).

For an example of some of these principles being violated, see figure 1. The distracting image of police cars in the background is a prime example of what Tufte would call “chart junk.” Also, the title is inaccurate, making the resulting data confusing; this chart does not track crime rates, but rather tracks the percentage of *change* of crime rates spanning two decades – a subtle but important difference. From figure 1, one could reasonably conclude that New York City had the lowest 2008 crime rate of all of these cities, which is not true. In figure 1 the crime rate for each city in 1990 was used as the benchmark (set to 100); then, based on the increase or decrease of crime in the ensuing 18 years, the line goes up or down. Of course, the crime rate of all of these cities was not equal in 1990, so this chart’s “lie factor” is apparent, as the images do not correspond with reality.

With the Tufte principles in mind and with an understanding of what can go wrong if we ignore them, a larger story can be told with the graphics created from our data. Four questions can guide those making the decision to create or evaluate graphic imagery to tell this story. These questions are as follows: (1) Which elements of the data should be highlighted? (2) How can these

elements be best represented? (3) How many dimensions of the data can/should be correlated? (4) Can users interact with the graphic? Answering these questions will allow the next steps of the process to organically materialize.

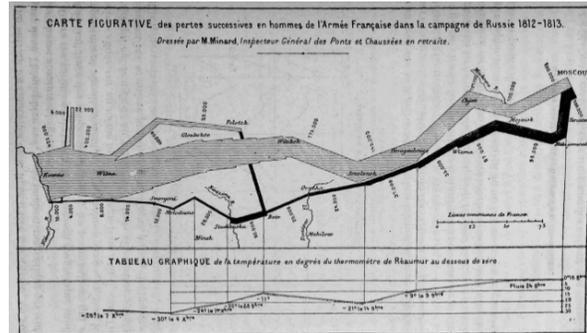


Figure 2. Napoleon’s March to Moscow (Marey, 1878).

Edward Tufte’s favorite example of combining perceptual dimensions is *Napoleon’s March to Moscow* by Charles Joseph Minard (Figure 2). This image highlights five distinct elements in a simple, uncluttered image—namely, the size of Napoleon’s army (the width of the streams), the direction in which they were marching (the color of the streams), temperature and time (line graph at bottom), and their geographical locations (position of the line at a given point). With a single static image, an entire narrative has been told, down to specific details. One potential improvement could be a user interaction element, which of course was not available in 1869 when this image was designed.

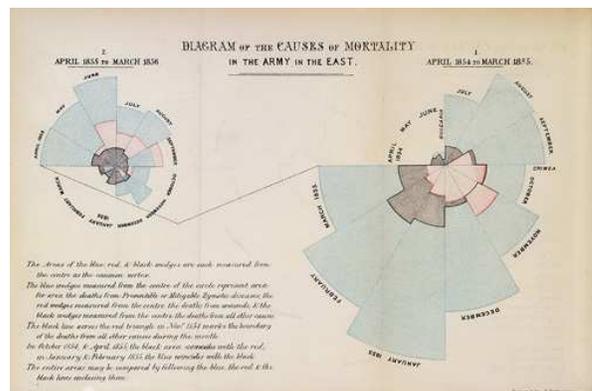


Figure 3. Diagram of the causes of mortality in the army in the east (Public Domain)

Florence Nightingale provides a second example of imagery being used for practical purposes, as she designed the above to petition for better health conditions for soldiers in the Crimean War. While presenting numbers and data are excellent ways to support an argument, visualizations like these allow one to show the data in a way that cannot be ignored. Nightingale's *Diagram of the causes of mortality in the army in the east* (Figure 3) showed that most of the British soldiers who died during the Crimean War died of sickness rather than battle wounds or other causes. It also showed that the death rate was significantly higher in the first year of the war, before Sanitary Commissioners arrived in March 1855 to improve hygiene in the camps and hospitals.

With this theoretical foundation established, it is an aim of the Center of Alcohol Studies Library to apply the practice of visualizing data with an ongoing research interest – namely, the quantification of scholarly reputation. As more data about scholarly reputation becomes available, many platforms have begun to use visualization tools to analyze and simplify the data for a wider audience. Reputation can be applied to individual scholars, their affiliated institutions, the journals in which they publish, or even their individual works. This paper will provide some analysis of tools that are used to display this concept, which is an abstract term that can be defined using several criteria.

Part 1 - How to define “scholarly reputation”?

Reputation is a fluid term that means a variety of things, depending on context. Bornmann and Marx (2012) provide what they call the “Anna Karenina Principle,” in which they posit that successful scientists are much like Tolstoy's idea of happy families in that they are all alike in having filled the three prerequisites of publication, citations, and new scientific discoveries. On the contrary, each unsuccessful scientist is *uniquely*

unsuccessful, in that they fall short of meeting one or more of these prerequisites in various ways. In order to establish a positive reputation, one must meet specific established and widely accepted criteria in terms of publishing, citations, and discoveries.

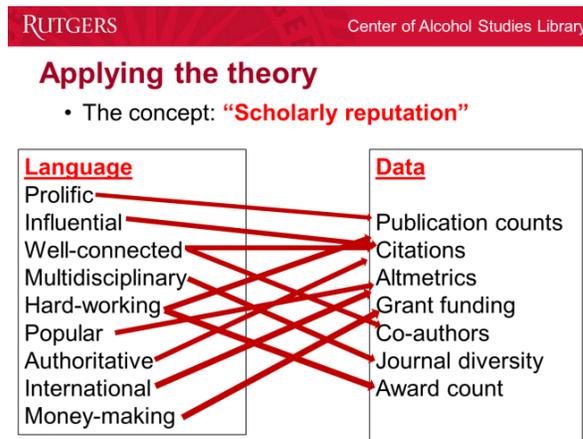


Figure 4. Scholarly reputation depicted using language and data

For the purposes of this paper, these three criteria will be combined, since landmark discoveries are assumed to be well-represented in terms of publishing and citations. In a general sense, this paper's definition of scholarly reputation is similar to Bornmann and Marx, in that it is defined as the output, impact, and activity of scholars, the venues in which they publish, and in their individual works. This paper purposefully broadens the Bornmann and Marx definition because part of understanding how reputation is best visualized is to understand how particular tools and applications *define* what is important. One tool might choose to focus exclusively on pure quantity of academic output, while another is more interested in the quality of that output, and a third focuses on the way the output is disseminated. None of these definitions are right or wrong, per se. Instead, they each focus on a particular aspect of academia, which is why this report is both a critical review as well as an idea board for one all-encompassing “scholarly reputation” tool that

selects and uses the best and most representative aspects of scholarship as it is understood in the field of information science and beyond.

Part 2 - Scholarly Impact (Traditional Measures)

a. Introduction

Measurement and quantification of academic production is tied to a few traditional measures. Typically, publication in an academic journal signifies some *new* contribution to whatever field one is publishing in, so perhaps the simplest way of measuring one's academic credentials is in terms of pure quantity of published research. If one researcher has 150 publications to another's 15, then simple math dictates that the former has been ten times more productive than the latter. The simplest visual accompaniment of this would be the length of the publication section on one's Curriculum Vitae, which can be found on web faculty directories, personal web pages, and in any number of employers' filing cabinets (round or otherwise). Slightly more sophisticated visualizations of this measure are usually

presented as a simple line graph or bar chart, such as in the case of Scopus, Web of Science, Microsoft Academic Search, and the browser add-on "Scholarometer," all of which show the year-by-year output of specific authors. Beyond simple tallying, documents can further be broken down, such as by affiliated institution, publication title, publication type, or subject area. Because of these extra breakdowns, some applications have several charts that show some more qualitative evaluations of a researcher's output. Color-coded pie charts (Scopus) and bar charts (Web of Science, Scholarometer) are typically used in these cases. Some applications are simply static charts that do not allow for any user interaction, while others offer details-on-demand when hovered over. Beyond that, there is not much more to visualizations showing total citations.

The next level of measuring academic impact is by tracking citation data. If a publication really is offering something new to the collective bank of knowledge, it will likely be useful in further research on a similar topic, and it would follow that the publication will be cited accordingly. Traditionally, citation analysis has been the preferred method of quantifying a researcher's influence. This leads to such measures as total citation count, citation average, the h-index (a metric developed in 2005 by Jorge Hirsch in which an author has published h articles with h number of citations) (Hirsch, 2005), the h_s -index (in which the h-index data are normalized for comparison across disciplines), and others. Luckily, citation data is abundant and has been compiled and analyzed since as early as the 1960s (Sher & Garfield, 1965).

b. Scopus Author Profiles

Scopus is a large database of citation information and article abstracts dating back to 1970. Using their large data set, they have created a built-in author evaluation tool that allows one to view an author's output in several respects. On the author's overview page, some of the main evaluative measures

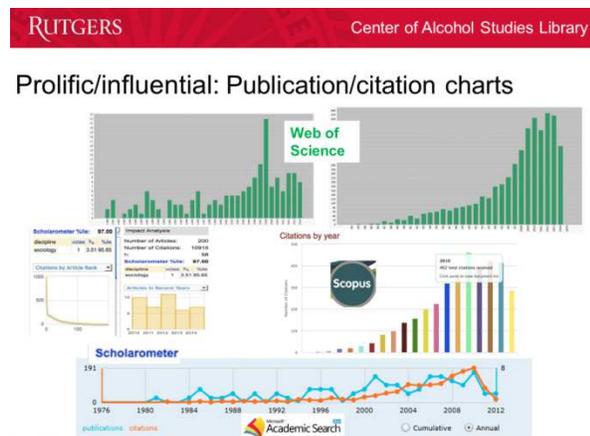


Figure 5. A sampling of traditional scholarly impact visualizations

are displayed in a table underneath the name and affiliation, including total documents, total citations, h-index, number of co-authors, and subject areas. On the right hand side is a bar graph with an overlaid line chart, which visualizes the amount of documents published and citations received per year. Because there are two charts showing potentially related, yet entirely distinct measures, there are two separate y-axes that measure the same length. The x-axis is a simple chronology of years. There is some user interaction involved in this chart, as all of the bars in the bar graph (representing a single year's worth of publications) and all of the plot points on the line chart (representing a single year's worth of citations) are hoverable, showing total numbers in a pop-up window, and clickable, which leads the user to a separate page listing either the documents published in that year, or a list of the articles that cited the author in that year.

Scopus offers further visualizations that delve a bit deeper into the author's total scholarly impact. A link to "Analyze Author Output" leads to a series of charts, one of which is a pie chart that breaks down the author's total work by the journal in which he or she published, with each segment of the pie tagged with a percentage, and when hovered over, leads to a pop-up information window that can be clicked for the list of publications. A similar chart breaks down the publications by subject area. Another is a donut-shaped pie chart that breaks down documents by type (e.g., journal article, editorial, conference paper, etc.), which offers the same information upon hover and click. Finally, the h-index is visualized in a unique way, showing a shaded descending line graph that depicts the total number of citations per document (again hoverable and clickable), ordered from most to least cited, with a line emerging from the nexus of the graph to intersect with the line chart at a point that represents the author's h-index. This is a great example of graphics being used to make complex metrics understandable using basic features.

[c. Microsoft Academic Search Author Page](#)

This application offers a simple double line chart that contrasts the number of an author's publications with the number of his or her citations. The chart uses color to distinguish the two lines (citations in orange, publications in blue), and a circular plot point on the individual lines mark each year. User interaction is very limited, as the graph is mostly static, save for the ability to switch between *Cumulative* and *Annual* radio buttons. The *Cumulative* button shows a chart with the y-axis ranging from zero to the total amount of citations the author has received. Because the line chart is cumulative, both the citation and publication lines are constantly increasing. Secondly, because citation counts tend to far outnumber the number of publications an author has, the line for the latter ends up being almost completely horizontal and clinging to the bottom of the chart, which could be seen as having a significant "lie factor" and misrepresenting one's output – though it is interesting to see when the two lines begin to uncouple and the citation counts begin to separate from publication counts, as this may be considered a crucial moment in a researcher's career. A slightly more representative chart is provided when clicking the *Annual* button, as it utilizes two separate y-axes, with citations marked on the left-hand side (ranging from zero to the most citations an author received in a single year) and publications on the right-hand side (ranging again from zero to the most publications produced in a single year), again using two color-coded line charts. Because of the dual y-axis ranges, comparisons can be more easily made between the years in which the researcher is more or less productive and years in which the researcher is cited more or less often. A second interactive feature is the ability to hover over any given plot point and see the exact number and year that has been plotted. A potential improvement in this case could be to make these plot points clickable, leading to further information, such as a list of publications or citations from that year. A

second criticism is the fact that charts end in the year 2012, indicating that this tool has since been discontinued.

d. Web of Science Core Collection Citation Report [Subscription only]

Web of Science is an enormous database of publication and citation information, spanning nearly all disciplinary fields of academia. The particular application focused on here is its Citation Report generator, in which one can create a report on a collection of publications. Theoretically, any group of publications can be collected and analyzed, but for our purposes, there is a somewhat roundabout way to focus on a single author. First, the user must search by the author's name. Because of the issues with potential ambiguity in names, the best way to zero in on the correct author is to select a known publication, and then click on the author's name on the ensuing article's record page, which will lead to something called *Distinct Author Record Sets*. This shows all of the potential names and affiliations that might match the correct author, from which the user can select one or more names. Once selected, the upper-right hand corner of the page offers a *Create Citation Report* link, which leads to two distinct bar graphs – one showing the number of published items in each year, and the other showing the number of citations in each year – next to a list of relevant statistics, such as the number of results (read: articles) found, the total number of citations (both with and without self-citations), the total number of citing articles (again, both with and without self-citations), the average citations per item, and the author's h-index. Unfortunately, outside of the ability to view a more expansive chart showing all years (by default, the charts display only the past 20 years), there is no user interaction available, as the graphs are simply static jpeg images.

Something unique about this tool is the ability to break the collected articles down by sixteen distinct categories using the *Analyze*

Results link. Some of these categories include Author, Country, Editor, Funding Agencies, Document Type, Affiliated Organization, and Research Area. The results are offered in a table with a sparkline-like (Tufte, 2006) bar chart that compares the percentages. This is a useful and interesting application, but much like in the case of the Citation Reports, the resulting display cannot be manipulated by the user.

e. Google Scholar Scholarometer

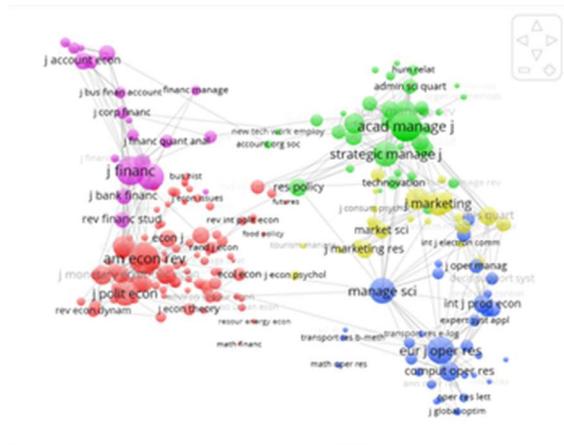
The Scholarometer is a plugin for the Firefox and Chrome browsers that allows a user to search for an author by Google Scholar ID, Name, Group, or with advanced filtering methods. To search by name, the user must tag at least one discipline that he thinks the researcher specializes in, in order to crowdsource relevant tags for individual authors. Once selected, an author's works are displayed in table format, organized by citation totals. On the upper-right hand side of the page, in another table, is the *Impact Analysis* of the author, with statistics for Number of Articles, Number of Citations, h-index, and Scholarometer percentile, as well as the hs index, for each discipline. Underneath this table is the only true visual display of information, showing a simple, static png file with a line chart offering total citations on the y-axis and article rank (from highest to lowest) on the x-axis, so every resulting chart will be a descending line at various depths. Also available via pull-down menu is a bar graph that shows the amount of articles the author has published in each of the past five years. Both of these visualizations are simple, supplementary items that can be found in more detail and with more potential for user manipulation in the other analyzed applications in this cluster. One other visualization offered by the Scholarometer that deals with traditional scholarly impact is the ability to compare the top authors in any given discipline by h-index or by hs-index, again using bar charts (length) to visualize the comparisons.

Part 3 – Scholarly Impact (Academic Networks)

a. Introduction

While summing up, averaging, and displaying citation data as specific metrics on traditional charts is a perfectly reasonable way to display one's scholarly impact, perhaps a more visually appealing way of displaying impact is with the use of academic *network mapping*, which is the most abundant type of all the tools that have been evaluated. Citation analysis, as mentioned, has been around for quite a while, and data is constantly being collected. Because the data set is large and ever-expanding, mapping citation or co-author chains can simplify and allow a user to see just how interconnected an individual scholar, a group of scholars, or individual works can be, or, conversely, who and what is more fringe and less connected to the general conversation. Typically, these networks are shown as a web of interconnected authors or works; but as we will see, there are several ways of displaying this data that can offer immediate and abundant information about how scholarly reputation is connected.

b. VOSviewer



Source: vosviewer.com

The VOSviewer is a highly sophisticated visualization tool that can be used to map not only author networks, but also those of disciplines, journals, individual articles, and even keywords. The approach taken in this application is to combine traditional mapping techniques with clustering techniques, which use color-coding, proximity, and size to organize the data points being presented (van Eck et al., 2010). The resulting examples of maps offer color-coded clusters that are hoverable but not clickable. A benefit of this tool is that it is not tied to any particular database like the Scopus/Web of Science/Microsoft Academic Search tools analyzed above, so theoretically any data can be imported and visualized, if cleaned up and presented properly. Another useful aspect is the ability to switch views, ranging from the default cluster model, to a heatmap (called “density view”), or scatterplot view. One drawback to this tool, based on some of the examples provided, is that a truly semantic zoom is missing; details are merely labeled on each plot point and displayed underneath the visualization and do not change as a user zooms in on the map. Further, it is difficult to understand what the color, the size of the plot points, or even the proximity of the plot points indicates, as each requires the user to look outside of the visualization for a key. This contributes to a non-intuitive experience, especially for the novice user. Outside of the ability to hover over plot points to see what each represents, and to zoom using the scroll wheel or navigational arrows, user manipulation is fairly limited.

c. CiteSpace

CiteSpace is a more focused networking tool that deals exclusively with citation mapping. Circles containing many multi-colored bands, with each color representing a specific year, represent an author's individual publications, with each band's width representing the number of times the work was cited in that year. Multi-colored tethers represent the year a work was first cited and

tie the individual circles together. As the network grows, clustering techniques are used to organize publications by theme, and the most prominent themes are labelled, with the size of the font corresponding to the size of the clustered theme. This tool is similar to the VOSviewer in that data can be imported into it, but it improves upon VOSviewer's less dynamic nature by allowing for more user manipulation, such as the ability to right-click a node and see its data history in greater detail, or allowing the user to switch to a timeline view as opposed to a clustered one (Chen, 2014).

d. vxInsight

The vxInsight uses a "landscape" visualization, which represents the activity of a scientific domain within the landscape's peaks and valleys. The higher the peak is on the graphic, the more active its representative domain. Boyack, Wylie, and Davidson (2001) used this tool to find potential competitors and partners in a corporate context. Yet this organizational method could easily be used to find the most active researchers within a specific academic discipline, using either citation analysis or co-authorship as a unit of measure. Similar to the previously analyzed tools, this tool allows for imported data and uses a clustering algorithm to bunch the like elements. Further, a time element is included, allowing the user to see how the valleys and peaks grow, retreat, or move, providing a visual demonstration of the abstract and conceptual idea of emerging trends in a given discipline. By adding depth and motion to the color, size, and proximity dimensions explored in the previous tools, this tool allows for a more easily understandable final product.

e. HistCite

HistCite was developed by Eugene Garfield, considered by many to be the father of citation analysis. His product offers a simple, clean view of how publications are

tied to one another using citation analysis. The individual publications are represented by circles, sized according to their Citation Score, calculated by number of times cited. The global citation score (GCS) is the overall number of citations, while the local citation score (LCS) is relative only to the collection being mapped. Each circle is given a unique number (the node number), which allows for easier identification and the ability to seek further information in the complementing table. Hoverability is supported via a mouseover that brings up abbreviated data about the selected publication. The y-axis is a timeline, and circles are connected by arrows, showing the direction of the citations. Also available is the ability to label nodes with author names to decrease abstraction using dynamic labeling.

f. CitNetExplorer

CitNetExplorer follows up on HistCite, by offering a more direct approach to showing author connections, representing publications as simple circles connected by lines on a 2-dimensional plane which, according to van Eck and Waltman (2014), "can handle much larger citation networks, possibly including millions of publications and citation relations" (p. 803). Created by the same team that designed the VOSviewer tool, key visualization elements are present, such as direct labelling, the ability to zoom and pan, and a specific focus on drilling down into a massive network to find more granular detail on specific sub-domains. Of particular interest to this paper is the ability to "study the publication oeuvre of a researcher" (van Eck & Waltman, 2014, p.805), by seeing the specific works he or she references at different points in his or her career, as well as the works that are in turn influenced by his or her work. Similar to the HistCite tool, a benefit of this visualization lies in its simplicity. Even a novice user can immediately understand the vertical timeline, the directly labelled plot points, and the clean connecting lines. Some manipulation is

apparent, as specific works can be “marked,” changing them from circles to squares, and networks can be narrowed or expanded, though not always within the visualization itself. This is a potential area for improvement, as presently the narrowing or expanding of a network is completed by changing maximum distance settings in a dialog box. Finally, clustering of the individual nodes is available, though again controlled via a dialog box, which complicates matters a bit.

g. Microsoft Academic Search

Microsoft Academic Search, already discussed in terms of its bibliometric aspects, also offers visualizations depicting author networks. Two types of network mapping are offered. First, from an author’s profile page, one can select the Co-author Path, which ties authors to one another based on works in which they were co-authors. All authors with a profile picture are represented by a circle containing that picture. A generic silhouette represents authors without a profile picture, but all nodes are dynamically labelled with researcher names. The tool provides the user a simple ability to click and drag faces around in order to isolate and analyze connections in whichever way the user may see fit. When the connecting lines are hovered upon, a number opens up, indicating how many publications the two authors worked on together. Clicking on this number, in turn, triggers a list of those publications. When the author nodes are clicked on, the graph dynamically changes to move said author to the center of the graph and to depict that author’s connections. The same principles are used in the Citation Graph, which can be accessed from individual article records in Academic Search, the main difference being that the nodes appearing on the map are the top 44 authors, based on total citations. The connections the user can view are the number (when hovered) and list (when clicked) of citing publications. Finally, there is the Paper Citation Graph, which adheres to the same principles as the first two, but at the article level.

h. Web of Science Citation Maps ***[Subscription only]***

Web of Science is the oldest and perhaps most trusted citation analysis software available. Given their incredibly large data set, a user can view a citation map of any article in its database in either the “cited” or “was cited by” categories for up to two generations in either direction. Each article is represented by a simple color-coded square with a single line tethering the root node to all of the others in a 3-dimensional fanned-out display. The colors of the squares do not seem to represent anything, but simply provide a contrast to the many other articles stacked among one another, so that each individually stands out. If there are too many articles to fit onto the dedicated space for the visualization, the articles collapse and only the tethered lines are shown, unless the user drags the graphic in such a way as to view them. The dragging and clicking inside of the chart is intuitive and can be used to isolate specific articles, which when hovered upon offer pop ups with their bibliographical information. Another useful feature is the time slider, which allows the user to see the rate of citations on a year-by-year basis. Because the start year is fixed, one can only see citations from the article’s publication year up to a chosen year and cannot select particular “slices” of time starting in the middle. Also, the time slider forces the user to click the *Re-create map* button for it to generate a map of the newly selected time span, which could be improved upon with a more fluid animation that would allow the user to see the changes in a more dynamic way. Another helpful aspect of this tool is the ability to click on one of the listed publications in the table below the visualization, which highlights the publication representation with a thick red border and allows the user to easily find it. When double-clicked, the full bibliographic information is displayed in the lower-right hand corner with a link to the record in Web of Science.

Part 4 – Total Impact

a. Introduction

While reputation has traditionally been represented by publications, citations, and related metrics, an emerging field has come into the fold in the area of what this paper calls “Total Impact” measures. These measures typically build on the pioneering work whose visualizations were analyzed in parts 1 and 2 of this paper, but expand further to include popular products, data sets, software, non-traditional dissemination methods, and other ways of tracking scientific impact. The first two tools analyzed, Altmetric and ImpactStory, collect outside data and present it in a way that offers a new perspective on scholarly output. The latter two tools, Research Gate RG Score and Academia.edu Analytics, are closer to social media profiles aimed at academics, each with their own measures and display functions to highlight individuals’ influence. The visualization tools include maps, graphs, charts, and badges, each offering unique usage of colors, position, and other perceptual coding tactics to represent an author or the work produced.

b. Altmetric

Altmetric is based on the belief or assumption that in our current world of constantly flowing digital information, traditional publications and formal citations form only one part of the much larger general conversation. Its second underlying assumption is that evaluation of articles should have less focus on the journal in which they are published, and more on the amount of attention they receive independent of said journal. Further, there is a push to move away from numbers that might produce a “false sense of precision” (Davis, 2013) and instead aim for a more qualitative evaluation. An example is the Journal Impact Factor metric, a journal-level metric discussed in greater detail below, and not considered to be granular enough to have any significant

bearing on any single article contained within. Instead, the Altmetric platform values dissemination in all its forms, including in tweets, shares in citation management software such as Mendeley, and references from popular journals and/or blogs. Its visual element comes in the form of an embeddable badge that an author can place on an article’s hosting page. This badge uses bands of color to proportionately display the types of shares an article receives, e.g., royal blue representing tweets, red representing news outlets, yellow representing science blogs, and so on. The number in the center is less transparent, deriving from a weighted ranking system of these shares and references based on the components “bias, promiscuity, and reach,” formulated by an algorithm that Altmetric has not disclosed (Davis, 2013). The badge’s use of color and proportionality is helpful to a user. Areas for improvement, however, include the capacity to hover for more detail on each of the colored bands, have the weights of the measures lie in the hands of the users to more accurately compare badge scores, and offer links to publication records or source data from which the badge and its metric was calculated.

c. ImpactStory

ImpactStory offers similar information as Altmetric, again utilizing shares from outside of scholarly circles. Its focus is on displaying the dissemination of both data sets and completed works published in open places such as Public Library of Science (PLOS), GitHub, and ArXiv. An author’s default profile page contains a list of selected works, with icons indicating impact, separated by whether those works are scholarly or popular. Scholarly works are presented in blue and include discussions of citations collected through Scopus, saves in Mendeley, and views in open networks such as in pdf format in PLOS or on figshare. Popular works are shown in green, indicating number of times discussed on twitter, viewed on

ImpactStory itself or in html format on its hosting page, and cited in Wikipedia. An addendum to each icon with a number of recent additions from the past week is also displayed. Each icon can be hovered over for detailed information about its sources, and clicked on to go to the publication's own ImpactStory record. A second visualization comes in the form of a geographical map with each country ranging from lighter to darker shades of blue, indicating an author's regional impact. When hovered over, the number of shares, views, and tweets is displayed in a pop up box, and when a country is clicked on, the publications with activity in that country are displayed in a list.

d. ResearchGate RG Score

ResearchGate is similar to ImpactStory in that it consists of individual profile pages and compiles its own site-centric data such as download counts and views. It displays scholarly reputation by combining an author's output with how others interact with that output, coupled with the statistics that ResearchGate keeps for its individual users. One of the metrics they use is called "RG Score," which is "a metric that measures scientific reputation based on how all of your research is received by your peers" (RG Score, 2014). By including contributions to one's profile, interactions with others, and negative results and data sets, ResearchGate aims to go beyond published work and instead use all scientific activity when calculating how a researcher earns a reputation. To visualize this, the RG Score page includes a pie chart of scientific activity, breaking down the author's output by publication, questions asked, questions answered, and number of followers by percentage of total activity. The green circle in the screenshot above is a somewhat poor example of this, as it is a chart that displays 100% publication. There is also a bar chart showing the author's RG Score over time, with details available upon hover, and a tile bar that compares the author's score against all of ResearchGate's members.

Finally, on the statistics page, line charts show recent view and download activities of the author's works.

e. Academia.edu Analytics

Academia.edu is another scholarly social media site with a focus on sharing research. Their theory is that with the move to digital distribution, scientific content becomes more easily measurable, and hence they launched an Analytics function in 2012. The key visuals are line charts of profile views, document views, and document downloads, along with donut-shaped pie charts showing where web traffic is coming from and which keywords led users to the author's profile. Much like ResearchGate, Academia.edu counts data sets, code, videos, and other media as equal to published journal articles, so this wider distribution of scientific knowledge is acknowledged. All charts are hoverable, giving precise information at each plot point. Finally, the Country View feature is reminiscent of ImpactStory, with lighter to deeper shades of colors on the map's individual countries representing the relative amount of activity in a geographical sense.

f. Kardashian Index

Finally, perhaps as a counter-measure to these more general metrics, there is the Kardashian Index, which measures the "discrepancy between a scientist's social media profile and publication record based on the direct comparison of numbers of citations and Twitter followers" (Hall, 2014). While a bit frivolous, its inclusion is offered to show that reputation at its essence is a matter of perspective and that one should view these visualizations with a critical eye, especially in regard to the data that is being used to create them. The visual is a prototype of a simple scatterplot graph that charts citation counts against number of Twitter followers and finds that typically the more highly cited authors tend to have a proportional number of followers. The outliers – those who have a

larger reputation than their output might dictate – are those whom Hall dubs “Science Kardashians” (Hall, 2014).

Part 5 – Journal Level Metrics

a. Introduction

To this point we have been concerned mostly with metrics at the researcher level or article level. It is also worthwhile to look at tools that measure reputation at the journal level, specifically in terms of Journal Impact Factor, the first prominently used bibliometric (Garfield, 1999), and its many descendants. Two tools—Journal Citation Reports and SCImago Journal & Country Rank—have been included in this cluster, and each provides information that can be used to compare journals in like fields. As might be expected, visualizations tend to be mostly charts and graphs.

b. Journal Citation Reports ***[Subscription]***

Journal Citation Reports, developed by ISI Web of Knowledge, offers an abundant amount of data on individual journals, such as Journal Impact Factor (average citations per article from the past two years), 5-Year Impact Factor, and Journal Cited Half-Life (how many years to account for 50% of the citations received by the journal). Most data is presented in tabular format; the few resulting visualizations are stacked bar graphs following a yearly chronology, with a shaded background depicting the Half-Life. These graphs offer a bit more information density than a typical bar graph, with the stacked colors differentiating between self-citations and other citations and a slightly lighter shade of blue indicating the data used for the journal’s current Impact Factor. There are separate graphs for citations *in* the journal and citations *of* the journal, with a third, more simplified bar graph showing the journal’s Impact Factor from the previous five years. Since none of these static, jpeg image graphs offer user manipulation, a potential

area for improvement would be to include hover effects, more details-on-demand, and a time slider to see how the metrics have changed over the years. Further, the user must navigate to each journal’s record to see these charts, so the ability to compare journals side-by-side is not available.

c. SCImago Journal & Country Rank ***(SJR)***

SJR offers several graphical displays of its data, including line charts comparing their proprietary metric SJR indicator to Web of Knowledge’s Journal Impact Factor and comparisons of external citations to self-citations at the journal and article-levels, and bar graphs displaying international collaboration, non-citable documents, and cited vs. uncited documents. On the whole, these graphics are much more varied and stronger than those offered by JCR.

As noted, the visualizations in these tools tend to be the tried-and-true stacked line charts and bar graphs, though they supplement charts with detailed explanations of each metric underneath each chart. Users can hover for the exact numbers at any given plot point (separated by year), or in the stacked charts/graphs they can hover over each individual color for the specific numbers in each portion of the stack. This is the extent of the manipulation allowed to users. Position, size, and color are the perceptual elements used, and data-ink ratio is high, without a lot of superfluous graphics or images. The lie factor seems to be low, with all graphs starting at zero and all measures seemingly proportional to the numbers they represent.

Part 6 – Grant Funding

a. Introduction

While publication and their dissemination seem to be the main criteria for scholarly reputation, grants are where much of the funding for these studies comes from, and hence a measure of one’s ability to secure

research grants can factor into their ability to ensure an institution remains funded to continue research into the future. There are two tools analyzed here that visualize this information, both dealing with government grants, since that information is open to the public.

b. Grantome

Grantome is a large database of grant funding information culled from several sources. Its visualizations are not readily apparent and one must search using Grantome's specialized search engine in order to see a set of results, underneath which charts and graphs are offered to aid in understanding the search results. This engine can be manipulated to offer visuals on any field that one searches in, including author, title, institution, study section, funder, grant type, and more. While researchers do not have individual profile pages, it is not difficult to search in a way to disambiguate a name from others (such as by providing an institution or topic in which that researcher specializes) and then analyze the results. The strength of this seemingly simple tool lies in its customization. For example, the y-axis offers a drop down menu from which the user can select total grants, total cost of grants, or average cost of grants. The x-axis can be assigned to funder, year, author, institution, grant type, institute, or study section. If the user would like two of these elements represented, there is a "Group by" option, which can color-code the data on any of these elements, and the user can select years, authors, institutions, etc. to compare to one another. Further, the data can be displayed as a bar chart, scatterplot, line chart, line with focus chart, or stacked area chart. Though the user is limited to using the data offered by its own database, thanks to these options variations between the input data (based on search terms) and output visualizations are essentially limitless.

c. NIH RePORTER

The National Institutes of Health provide a comprehensive database of all of its grants, which, according to its website, satisfies a "legislative mandate included in the NIH Reform Act of 2006 to provide the public with an electronic system to search NIH research projects using a variety of codes, including public health area of interest, and provide information on publications and patents resulting from NIH-funded research" (NIH RePORTeR, 2014). Much like in the case of Grantome, its visual displays are based on the data within its search results. Given its highly faceted query form, which can be broken down by name, organization, year, department, type, or any number of project details, the results can again be customized at a highly granular level from the results page in a tab called "Data & Visualize." From this tab, there are two sub-headings, one of which provides traditional charts, with columns, bars, and pie charts available. These can be summarized by administering institute, year, activity code, state, and more. Further, the charts can be grouped by project, funding, or project publications. While the user cannot manipulate the charts themselves, a table is provided with links that lead to lists of detailed records that have been compiled to form the data in each. While these charts are helpful, a more interactive chart is available via the "Circles" subheading, still in its beta phase. This helpful visualization offers a subdivided pie chart showing a hierarchy of the data and allowing for many aspects to be represented within a single image. Each slice can be hovered over for more information, clicked on to highlight its topic in the center of the circle, and double-clicked to have that portion blown up disproportionately to view the details of its subordinate features. Finally, there is a mapping feature, which shows locations of projects on a dynamic Google Map with all of its usual methods for manipulation and from which further information can be found by clicking on the map's plot points.

Unfortunately, the NIHMaps function, which provided a color-coded, cluster-based visualization of grants similar to some of the mapping tools analyzed in Part 3, is no longer in service.

Part 7 – Toward an All-Encompassing Prototype

Having spent much time analyzing and interacting with these visualization tools, it seems to us that an all-encompassing tool would need to be incredibly diverse, utilizing a wide breadth of data, visual techniques, and interactivity. Because the definition of scholarly reputation is so general and the

importance of its individual components can vary from user to user, the key to a useful tool lies in its customization. Secondly, because these metrics can become rather complicated even for those well versed in the subject, the ability to express these complex measures in natural language for the user is paramount. A third observation is that since data comes from several sources, there must be a way to either normalize data or give the user the data source option. Finally, perhaps the greatest challenge will be to display such abundant and complex information in a simple and clean way the user can easily manipulate and understand.

References

- Bornmann, L., & Marx, W. (2012). The Anna Karenina principle: A way of thinking about success in science. *Journal of the American Society for Information Science and Technology*, 63, 2037–2051. doi: 10.1002/asi.22661
- Boyack, K. W., Wylie, B. N., & Davidson, G. S. (2002). Domain visualization using VxInsight® for science and technology management. *Journal of the American Society for Information Science and Technology*, 53(9), 764-774.
- Chen, C., Ibekwe-SanJuan, F., & Hou, J. (2010). The structure and dynamics of co-citation clusters: A multiple-perspective co-citation analysis. *Journal of the American Society for Information Science and Technology*, 61(7), 1386-1409.
- Chen, C. (2014). *The CiteSpace Manual* (Version 0.65). Retrieved from <http://cluster.ischool.drexel.edu/~cchen/citespace/CiteSpaceManual.pdf>
- Davis, P. (2013, September 30). Visualizing article performance — Altmetrics searches for appropriate display. *The Scholarly Kitchen*. Retrieved from <http://scholarlykitchen.sspnet.org/2013/09/30/visualizing-article-performance-altmetrics-searches-for-appropriate-display/>
- Garfield, E. (1999). Journal impact factor: A brief review. *Canadian Medical Association Journal*, 161(8), 979-980.
- Hall, N. (2014). The Kardashian index: A measure of discrepant social media profile for scientists. *Genome Biology*, 15(7), 424.
- Hirsch, J. E. (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), 16569-16572.
- Imagine a pie chart stomping on an infographic forever. (2010, May 10). *Smashing Magazine*. Retrieved from <https://www.smashingmagazine.com/2010/05/imagine-a-pie-chart-stomping-on-an-infographic-forever/>
- Marey, E.-J., & Burndy Library. (1878). La méthode graphique dans les sciences expérimentales et particulièrement en physiologie et en médecine: par E. J. Marey. Paris: G. Masson. Retrieved from <http://tinyurl.galegroup.com/tinyurl/oAoZX>
- NIH RePORTER. (2014). Retrieved from <http://projectreporter.nih.gov/reporter.cfm>
- Piwowar, H. (2013). Altmetrics: Value all research products. *Nature*, 493(7431), 159.
- RG Score. (2014). Retrieved from <https://www.researchgate.net/publicprofile.RGScoreFAQ.html>
- Sher, I. H., & Garfield, E. (1965). New tools for improving and evaluating the effectiveness of research. In M.C. Yovits, D.M. Gilford, R.H. Wilcox, E. Staveley and H.D. Lemer (Eds.), *Research program effectiveness*. Proceedings of the conference sponsored by the Office of Naval Research (pp.135-146). New York: Gordon and Breach. Retrieved from

<http://www.garfield.library.upenn.edu/papers/onrpaper.html>

- Tufte, E. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.
- Tufte, E. (2006). *Beautiful evidence*. Cheshire, CT: Graphics Press.
- van Eck, N. J., Waltman, L., Dekker, R., & van den Berg, J. (2010). A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. *Journal of the American Society for Information Science and Technology*, 61(12), 2405-2416.
- van Eck, N.J., & Waltman, L. (2014). CitNetExplorer: A new software tool for analyzing and visualizing citation networks. *Journal of Informetrics*, 8(4), 802-823.
- Waltman, L., Van Eck, N. J., & Noyons, E. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4(4), 629-635.

Contact the authors

William Bejarano, MLIS, MLER
Senior Information Specialist,
Rutgers Center of Alcohol Studies
607 Allison Rd.
Piscataway, NJ 08854 USA
Phone: 848-445-4442
bejarano@rutgers.edu

Judit Ward, PhD, MLIS
Director of Information Services,
Center of Alcohol Studies,
Rutgers, The State University of New Jersey
607 Allison Road
Piscataway, NJ 08854-8001 USA
Phone: 848-445-4442
jhajnal@rci.rutgers.edu