Visualizing researchers' scientific contributions with radar plot

Manh-Toan Ho[1]

Abstract

The essay advocates for diverse approaches in presenting a researcher's scientific contributions in a project. Taking inspiration from sports journalism and its visualization of football players' data, the essay suggests that a radar plot, incorporating CRediT contributor role data, enables multiple authors of a scientific paper to illustrate their contributions in a more specific manner. The suggested method, though subject to bias reporting, pays credit to different aspects of a research project, from conceptualization, analysis, administration, to writing and revising. It not only enables both academics and lay readers to better understand the considerable amount of work required in every project but also calls for the need to employ diverse viewpoints in science.

Introduction

Scientific profiles of researchers use single-number metrics, such as publications, citations, or h-index, to present an overview of the authors' impacts. Online scientific profiles, such as the Web of Science or ORCID, primarily utilize single-number metrics for visualization. For example, the Web of Science provides a plot showing times cited and publications over time on a researcher's profile. While it can be argued that single-number metrics offer a simple yet elegant portrayal of impacts, the glaring problems of metric manipulation and substituting metrics for actual quality are damaging to the scientific community [1,2]. Drawing inspiration from football analytics and data visualization, this essay aims to explore a different method to visualize a researcher's scientific contributions.

Why football analytics? In recent years, statistical analysis has indeed become an integral part of performance analysis in various sports, such as basketball and baseball. The book Moneyball: The Art of Winning an Unfair Game, and its movie Moneyball (2011), has popularized the concept of using statistical analysis to gain an advantage in sports, particularly for teams with lesser financial power. However, the success of statistics in basketball or baseball does not translate well to football due to the low-scoring nature of the game [3]. Hence, the scoreline does not necessarily reflect the performance of the team or a player. Moreover, as a data point often registers an on-ball

---

[1] Centre for Interdisciplinary Social Research, Phenikaa University; manhtoan212@gmail.com

action, off-ball actions are often neglected [4]. Despite the shortcomings of football data, football writers have employed different methods and visualization techniques to demonstrate a player's contribution. The visualization used in this article is the radar chart.

Plotting the radar chart with CRediT data

The radar chart, more commonly referred to as a pizza chart, is extensively utilized in football analytics [5]. This chart offers a comprehensive way to summarize various statistics of a player during different phases of actions in football—be it attacking, defending, or possession. Here is an example of a pizza chart:
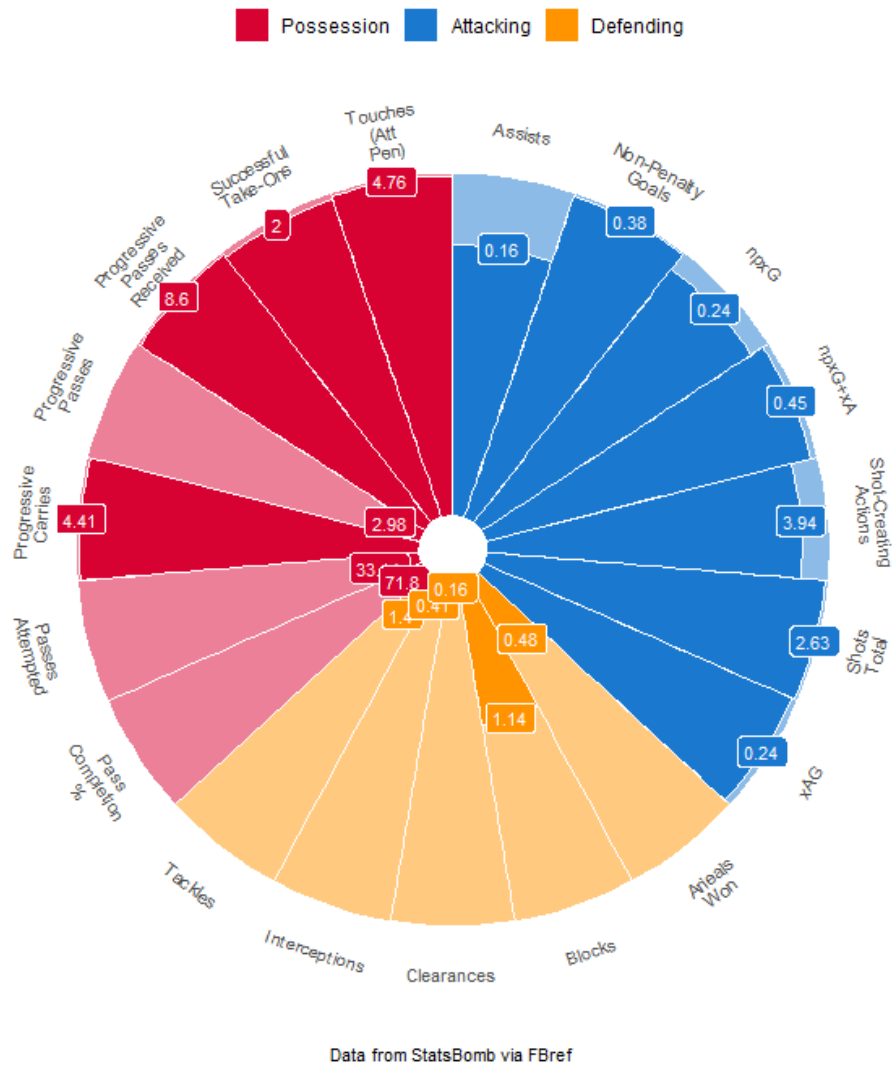

Figure 1. A pizza chart for the player Takefusa Kubo, created with R code provided in [6]

In science, while there is no consensus regarding research phases, the CRediT – Contributor Roles Taxonomy — is a common method for researchers to declare their contributions in a paper [7]. Presently, numerous publishers have adopted CRediT, including Cell Press, De Gruyter Open, Elsevier, SAGE, Springer, and others [7]. Additionally, free services such as Rescognito (https://rescognito.com/) also utilize CRediT to enable researchers to acknowledge their contributions and those of their peers. Rescognito is also using CRediT data for visualization, though the visualization tools are still in the beta stage. CRediT's 14 contributor roles are clearly defined and easy to understand. However, beyond a declaration in the manuscript, there seems to be no apparent utilization of CRediT.

Hence, in lieu of football statistics, this essay attempts to visualize researcher statistics using the 14 contributor roles from CRediT. These roles were grouped into 4 categories representing phases of research: Conceptualization, Analysis, Administration, and Writing.

Table 1: The 14 contributor roles and their groups (Adapted from [7])

Roles (Group)	Explanation
Conceptualization (Conceptualization)	Ideas; formulation or evolution of overarching research goals and aims.
Data curation (Analysis)	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later re-use.
Formal analysis (Analysis)	Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data.
Methodology (Analysis)	Development or design of methodology; creation of models.
Investigation (Analysis)	Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.
Software (Analysis)	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.
Project administration (Administration)	Management and coordination responsibility for the research activity planning and execution.
Funding acquisition (Administration)	Acquisition of the financial support for the project leading to this publication.
Resources (Administration)	Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools.
Supervision (Administration)	Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.

Validation (Administration)	Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs.
Visualization (Writing)	Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.
Writing – original draft (Writing)	Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).
Writing – review & editing (Writing)	Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages.

Figure 2 provides a visualization of CRediT roles for an early career social scientist who published 31 articles from 2018 to 2023 (please note that *Writing – original draft, Writing – review & editing, Conceptualization* were shortened respectively changed to *Writing, Revising, Concept* for clarity).
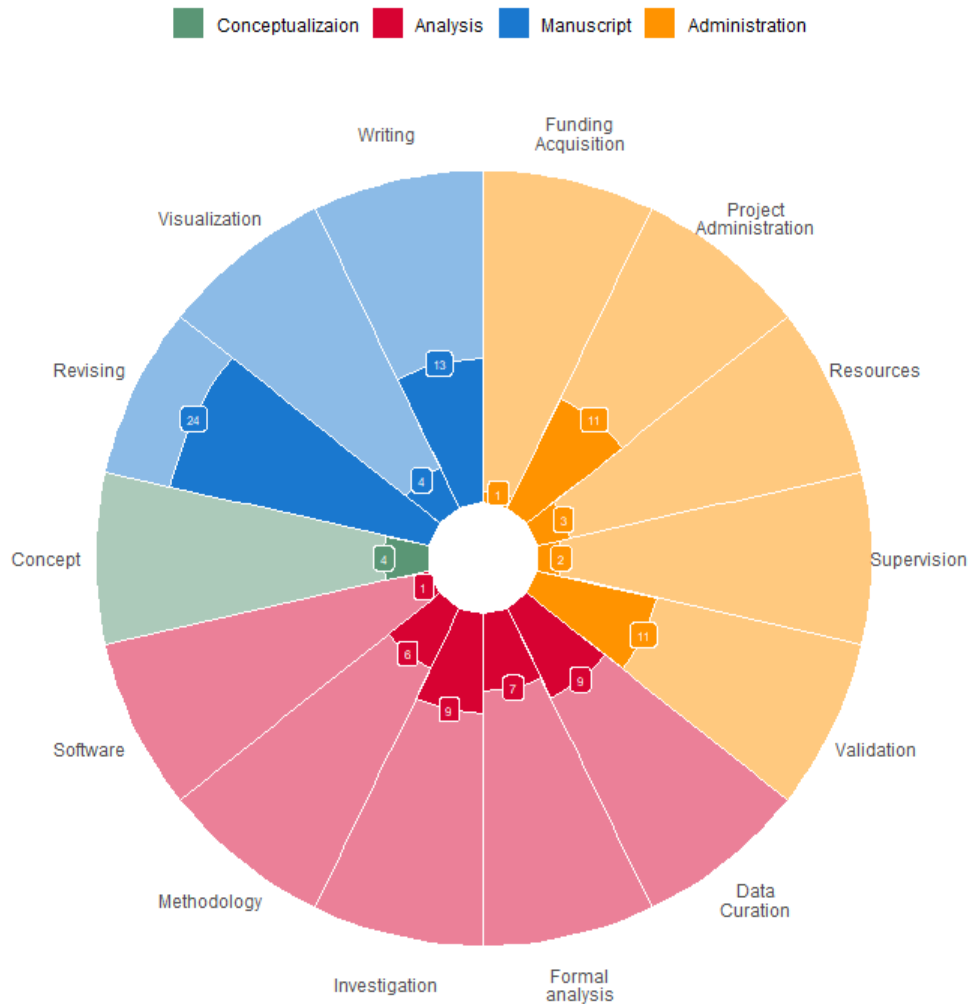

Figure 2: Researcher A's contributions from 2018 to 2023.

As depicted in Figure 2, Researcher A's primary contributions lie in writing and revising manuscripts, as well as responsibilities in data curation and investigation. To further illustrate the potential of the pizza chart, I have visualized a comparison between Researcher A and a hypothetical senior Researcher B in In Figure 3 (for visual clarity, I shorten *Project Administration* to *Ad*min, and *Funding Acquisition* to *Funding*). As you can see, Researcher B assumes the role of the senior researcher and makes significant contributions to supervision and conceptualization in the studies.
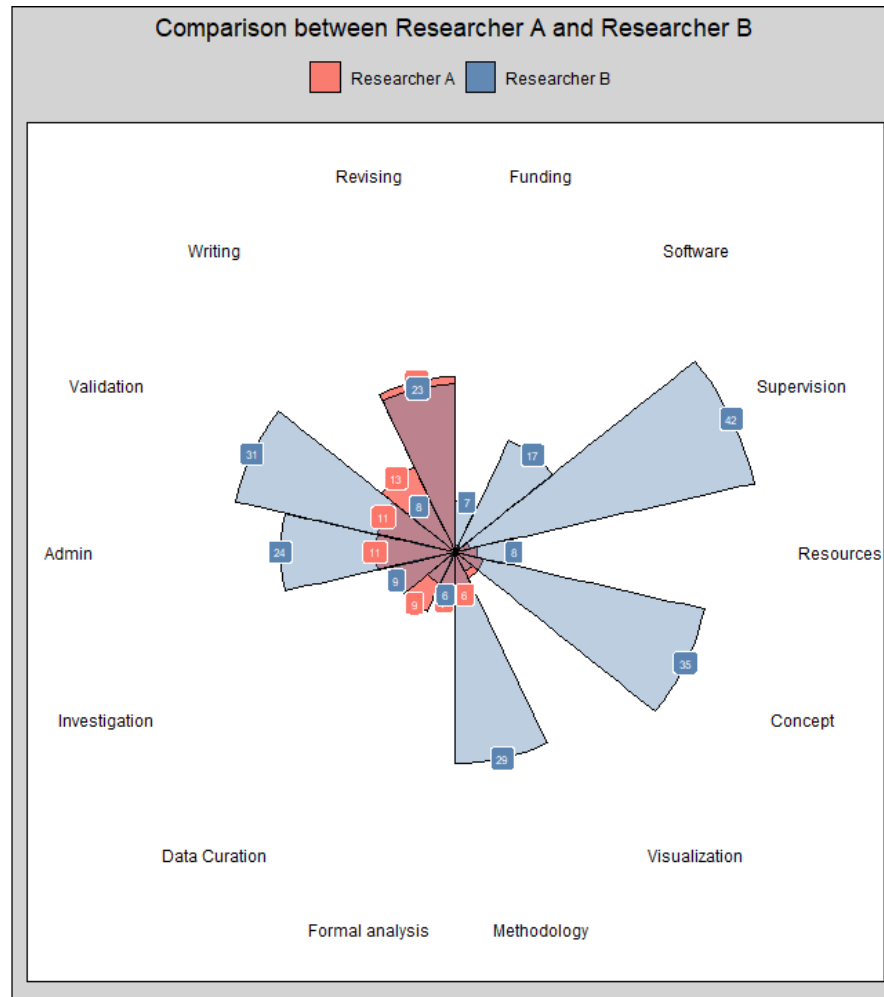

Figure 3: Comparing researchers.

A pizza chart can enhance the presentation of several aspects. Firstly, it effectively illustrates an individual scientist's contribution to a paper. This feature is particularly valuable for researchers in large research groups or fields where collaborative, multi-authored papers are common. Additionally, the chart serves as a useful tool for visualizing how scientific contributions evolve over time. For example, as a young social scientist, the focus was primarily on manuscripts and data (like Researcher A). In contrast, senior researchers tend to shift their contributions towards the administrative

phase or key aspects of the study, such as the main concept or methodology (like Researcher B). Variations of the pizza chart allow for comparisons between researchers and, with enough data, between a researcher and the field's average. Moreover, the chart can also represent author positions (first, corresponding, or co-author) and replace the researchers' images in the inner circle.

Limitations and Closing Remarks

Indeed, there is no one-size-fits-all method to fully capture the beauty and complexity of research or a person's career [8]. Therefore, we shouldn't limit our presentations of science to single-number metrics or only a handful of visualizations based on those metrics. I argue for diversity in presenting scientific contributions, providing examples like the radar chart and CRediT contributor roles data.

For a complex activity like science, it is quite peculiar to observe the limited ways scientists publicly present their work. The development of scientific journals and their ecosystem can be attributed to both historical and practical reasons. However, as Adam Mastroianni argued, the current ecosystem has its limitations and, in a certain sense, has failed us [9]. Even though the journal impact factor was not initially intended for rating science, the precise number has failed to facilitate a healthy system, despite its mathematical appeal [10]. In "The Seduction of Clarity," philosopher C. Thi Nguyen argued against the illusion of mathematical beauty. Using the standardized value systems of bureaucracies as an example of how clarity can exploit our cognitive vulnerabilities, he pointed out: "Quantified systems are, by design, highly usable and easily manipulable. They provide a powerful experience of cognitive facilities. It is much easier to do things with grades and rubrics than it is with qualitative descriptions" [11]. Similarly, single, easy-to-use, easy-to-understand metrics like journal impact factors, citations, or publication numbers make things simpler for everyone. However, in our pursuit of the clarity these numbers provide, the scientific community often sacrifices a qualitative, nuanced, and complex understanding of science.

In football, there exists a clear cognitive dissonance between what one observes on the field and the actual results. When Brazil experienced a humiliating defeat on their home turf during the 2014 World Cup against Germany, the underlying data presented a different picture, with Brazil actually superior in several aspects (shots, shots on target, possession, dangerous attacks, to name a few) [12]. Yet, Brazil lost 1 – 7 to Germany in that semifinal. This disparity in football necessitates that data analysts exercise creativity and caution when interpreting football data. Likewise, we can adopt diverse perspectives to discuss and understand science.

This essay advocates for diversity in the presentation of scientific contributions. The once favored single, easy-to-use, and easy-to-understand metrics have had their moments. It's time to move away from relying solely on these metrics and begin developing new languages to represent ourselves. My suggestion is to visualize the CRediT data. However, this is just one approach; there are other possibilities. Currently,

the availability of CRediT data poses a significant challenge in implementing this method. Publishers like Scopus, Web of Science, and others may need to consider providing it to the public [13]. Additionally, we should be mindful of potential biases arising from the self-reported nature of the CRediT data when discussing the results.

The scientific community has been calling for changes in how we use metrics. However, finding a suitable replacement remains a challenge. This article presents an attempt to adopt a type of chart from football analytics for research purposes. This marks an early effort, and there are still many shortcomings. Nonetheless, as science promotes increased openness and transparency, a new strategy is imperative for a cost-effective and more informative outcome [14,15].

Supplementary Information: R code used to generate figures is available in SI.

Gardener Comments

Anonymous 1 (9 years of experience as journal editor):
I like the approach. However, I would like to see more examples from other researchers. I suggest you pick a top researcher with a large number of publications, so one can draw a plot based on many data points. If you picked two researchers, you can compute a contrast version of the plot where you subtract one's scores from the other, showing their relative preferences for roles. Ideally, there would be a website where users can look up any researcher (e.g. by ORCID id) and see their plot, as well as the contrast with another researcher.

Manh-Toan Ho: Thank you for your comment. I have provided a comparison plot between two researchers.

Andrew Neff (Psychology PhD):
I agree that this visualization and the CRediT system in general provide a more accurate metric of experience than authorship alone, I think it could improve hiring and retention. I also agree a lot depends on the availability of CRediT data, and wonder if a platform like Google scholar would allow scientists to add this info post-publication if a journal doesn't naturally list it. I also am starting to wonder whether quantifying each CRediT category could also be helpful, and possibly even necessary, if some (lazier) people are going to use this info in hiring.

DAhmad Ozair (MD, clinical researcher with experience in bibliometrics and scientometrics research):

The article presents an interesting method of visualizing scientific contribution of authors to peer-reviewed publications. Certain considerations, however, limit the rigor of the work and of its external generalizability:

1. The author's own radar plot is drawn from 15 papers - which is a very small sample. They should strongly consider expanding this to papers beyond 2020, as this is not too challenging.

Manh-Toan Ho: I have updated the data, as well as adding one comparison plot.

2. Author says that "For a complex activity such as science, it is rather strange to see the limited ways scientists present their works." - saying this seemingly negates the entire field of bibliometrics and scientometrics which have extensive visualization options developed over decades.

Manh-Toan Ho: "*For a complex activity like science, it is quite peculiar to observe the limited ways scientists publicly present their work.*"

I have added the word publicly to clarify what I want to say in this sentence. My point is that the applications of "entire field of bibliometrics and scientometrics which have extensive visualization options developed over decades" are somewhat limited to a niche area rather than a widely adopted standard. When we describe someone's research career in public, we typically only use numbers of publications, h-index or JIF to illustrate. That is something I find strange.

3. One limitation that may be added is the lack of rigor in the CREDIT listing itself by researchers. In many labs and groups, CREDIT is an afterthought and authorship order continues to be the primary point of discussion. Additionally, despite efforts, the majority of papers published lack complete CREDIT-appropriate listing, which limits automatic analyses.

Manh-Toan Ho: I added this as a limitation to the discussion. However, this is a general problem when a new standard is being introduced, I think as many publishers are using CREDIT, it will soon be the norm and authors will have to take it more seriously. Moreover, I think the effort to use the CREDIT data will also improve the general perception of CREDIT as well.

4. Please consider adding to the discussion, the work being pioneered by the RESCOGNITO team. They have a great visualization tool using CREDIT already.

Manh-Toan Ho: I have mentioned RESCOGNITO in the article.

Josh Randall:

This manuscript describes an interesting means of visualizing researchers' contributions to the creation of publications. It takes advantage of the use of statistics in sports

analysis to develop a system for showing how authors can contribute in different ways for the final publication. Two criticisms that I have are: a different form of visualization does not fundamentally move away from the publication focused description of a researcher's merit and so many of the components of constructing a specific paper such as administration in a specific lab group may not translate well in any other setting. The most useful aspect of this visualization would be seeing how one specific researcher changes through time, but I do not think this solves the fundamental problems with using H index to determine employability or success of a researcher.

Manh-Toan Ho: Thank you for your comments.

I think your concerns are with the systematic ways of doing things in academia, and of course, that problems can't be solved with just using a new way of visualization. Scientific communities are constantly finding new ways, such as *Seed of Science,* so I think the accumulation of all the efforts will eventually lead to meaningful changes.

Duy Nghia Pham:
1. I think the "Software" role should be placed in Analysis, not Administration.

Manh-Toan Ho: Thank you for your suggestion. I have switched Software to Analysis accordingly.

2. In figure 2, all 14 roles should be visualized (even if some are zeros) for a complete picture of one's contribution and easy comparison with other researchers.

Manh-Toan Ho: I have updated the data and visualized all roles.

3. While the Credit data might be missing, the author positions (first, last, co- added up to 100%) are available on Scopus or WoS, which provide similar information on the contribution. These 3 roles can be plotted in radar chart too. The inner circle (instead of showing the footballer/researcher face) may be used for the percentage of corresponding role (e.g, as a pie chart).

Manh-Toan Ho: I have included this point in suggestions for future directions.

Vamsi Makineni:
Although the suggestion is interesting, I believe the author should expand upon the limitations and how we can solve them. The idea of a pizza chart for science can easily be ill-defined as the components are rough estimates of contribution. Also, one's role can be deduced in the acknowledgement section. It is apparent that early-stage researchers will mostly help with data collection and gradually transition to conceptualization as they gain experience so I do not see how the chart will provide very important information. Who should it benefit? At most, I see this proposal benefiting talented early-stage researchers who help in conceptualization. Of course, unlike football there is no easy point system to assign for this criteria. A single idea could be momentous to solve a problem the research team has had, so how many points should

this idea receive? I believe that the author should take these comments into consideration and rewrite the paper.

Manh-Toan Ho: My initial idea and preliminary results are simple: visualizing a researcher, and comparing two researchers. However, with enough data, I think there are more ways to show how a researcher 'performs'. For instance, comparing how a researcher contributes in a year to his team's average, or to the average of the authors with the same age. I think the whole community can benefit, with large enough data of course. Back to my comparison with football or sports, the rise of data analytics in sports eventually brought sports into the modern age with new technologies in recruitment or managing players' performances. So while I understand your skepticism, I think it is better to look at the bright side.

Dr. Payal B. Joshi:
The idea of using pizza chart/radar plots and utilizing solitary publishing record that features CRediT information is flawed. It is good to note that the author realized that when concluding as its quoted, "one size-does not fit all" however, going ahead with the idea seems odd or depicts a dilemma for a researcher. Do we really need pictorial representation of each scientists/researchers' output? Any metric or chart that is developed on a flawed premise is not useful. There is a constant effort that universities do not push metrics to quantify science and research integrity is already at stake. CRediT information is provided by the authors which itself is highly biased data. Author has completely missed the gift authorship and increasing disparities in today's collaborative papers. Most of the authors are listed to act as funders for open access fee funding.

Even if we keep research integrity aside, the idea is flawed. Why? - The answer is by the time you make a pizza chart based on CRediT from publishing record of past 10 years, and unfortunately few papers face retractions - what will be the case then? Is not the author's record questionable in the first place. If author has manipulated data due to which he/she faced retraction, how credible is the CRediT information? Has the present author thought about that and thus can accordingly plan the R code run. I urge the author to read literature of Vuong QH whose work is already cited in the present article known for his research integrity work.

My final take is rather than ideating on novel author charts, we need to think on ways to curb any new formulation of metrics as that makes us accountants and not scientists. We need to perform qualitative analysis and enhance post-publication discourse to truly attract talent and not manipulators. I, thus, do not recommend publishing this article, however, can be considered as a blog on Seeds of Science. Pardon me as I cannot be more positive for the present article.

Manh-Toan Ho: CRediT information, despite its flaws, is still one of the most orthodox and systematic information that publishers have on author contributions. So, my point is to find a better use of the information that we have, and already gathered. Also, I agree

that we need better qualitative analysis and enhance post-publication discourse, however, it doesn't mean that we should only focus on this aspect. Improving the way we talk about ourselves is also a way to move forward. Moreover, I think a lot of your questions are technical in nature, like the gardens of forking paths. There will eventually be so many micro-decisions to make when dealing with data, and I think the best way to deal with that is be clear of how you use your data and your methods.

Dr. Christian Thurn:
I like the idea and there are many great things in the article. The sentence that I enjoyed most was "For a complex activity such as science, it is rather strange to see the limited ways scientists present their works."

Please make an R-package or share the R code for the radar chart, so that others can use it as well.

Manh-Toan Ho: I have provided snippets of code, as well as making the R file available in the supplementary information.

I don't understand the second sentence: "The visualization of the single number metrics on online scientific profile like Web of Science is also minimal with citation or publications trends over time." Something weird with this sentence: "The radar chart (or more commonly, pizza chart) that has been widely used in football analytics [5]."

All else is good – short and clear

Manh-Toan Ho: Thank you for your comments, I have revised the sentences for clarity.

Jay Matthews:
In general I'm excited about the prospect of visually depicting skills, perhaps as an alternative to CVs, and finding alternative ways of showcasing competency—especially as careers become more diverse and interdisciplinary. Should this paper be published, I would hope that it could serve as a starting point for what I think might be a more important discussion, that is the validity of these measures. I believe you've covered this well in your Limitations and Conclusions section. In short, a visual representation that presents a broader view of contribution and participation could be beneficial and seems like it has a place in the future, however, this paper feels a bit more like a first draft. The

Difficulty of Measurement:

I started working on a project to develop measures (for which you've used CRediT data) in order to build a ranking system, but it quickly went awry as human measures have been known to do. The actual measurement of skills, competencies, and experience is interesting to me and could perhaps add value to the recruitment industry, especially in light of the 'academic arms race'. Measures that will accurately reflect not only participation and completion, but be an indicator for quality and skill is perhaps somewhere this can go? Not to mention the kinds of patterns we might find when

comparing the frequency of roles fulfilled and demographic factors like age, gender, field.

Comparisons to other work:

Perhaps a more holistic picture of the person could build on Dan Robles's Curiosume? Perhaps the same skills can stand and a dimensionality of skill level could be added? (e.g. my level of experience (or skill) vs how interested I am in it). DataCamp has a similar visualisation for all the platform members along core skills related to different career choices (Data Analyst vs Data Scientist). The 80,000 hours site has a lot of preliminary work for competencies and skill outlines for different professions too, if this were to be compared or used elsewhere. If nothing else, this idea could benefit from comparison to more than just football, though it is a nice and easy starting point reference!

Potential concerns:

- Why is the CRediT data so scarce? What might be preventing people from using this?

Manh-Toan Ho: Thank you for your above discussion, they are insightful and valuable. Regarding the scarcity of CRediT data, even though declaration of contributions is a common practice for most publications nowadays, I do not know many services that publicly declare to use this data for further analysis, or any well-known publications/reports that focus especially on CRediT data. There is Rescognito (https://rescognito.com/), which makes use of CRediT data. However, their system is based on what authors claim, not an automatic data scraper. This makes it hard for any application of CRediT data on a mass scale.

- Might there be negative effects on researchers' careers for making this more explicit? (e.g. If I eagerly participate in project to grow my career but not given opportunity to do more than admin, might this exclude me from future non-administrative opportunities)

Manh-Toan Ho: The roles in CRediT are industry standard. At a certain point in a researcher's career, he/she will contribute in that role. So I do not think there will be any negative effects.

- Is this CRediT system useful and, if so, widely applicable across different types of research? Does standardisation increase commodification?

Manh-Toan Ho: So far, I decided to use CRediT because they are the standards for most publishers, as I have written in the article. My idea is when most publishers actively collect this information, it would be easier to aggregate, collect and apply the data on a large scale.

- Productivity is a difficult thing to measure, especially when producing knowledge work. In football (and perhaps even in other jobs that are more clerical or unit-based), there is a very clear measure or success with ROI attached to it. For research, it'd be difficult to

think of a good comparison. Knowledge work doesn't always have a direct correlation between time or training and the quality of output. Perhaps this will change! I appreciate the note in the limitations about how individual stats don't necessarily correlate with winning, however, there are enough knowns to build a prediction market around football outcomes. Is the same true for research? And if so, what would the predictions be of? What is the ROI of a single research paper, and can it ever truly be counted by itself, and in how many years of increments? This is not meant as a criticism of this paper or idea, but rather as an encouragement — because researchers are dependent on jobs, organisational rewards systems and performance reviews to make a living, I believe there may be a lot of value in taking this implicit measures and modelling them more explicitly because it may enable us to interrogate the methodology and assumptions more incisively!

If all goes well, this practice could create great opportunities to identify and take on career development projects in areas you'd like to present more strongly (for instance the author of this paper on visualisation has a 2 on the visualisation spoke of the graph).

Étienne FD:
Although it would be beneficial to refine the idea further, radar charts for scientists' contributions are a highly interesting idea to explore. It's valuable to spend some time researching and communicating how scientists actually perform their work and how that can affect the way science is conducted.

References

1. Tran T, et al. (2020). Scrambling for higher metrics in the Journal Impact Factor bubble period: a real-world problem in science management and its implications. *Problems and Perspectives in Management, 18*(1), 48-56.

2. Vuong QH. (2019). Breaking barriers in publishing demands a proactive attitude. *Nature Human Behaviour, 3*(10), 1034.

3. Stevenson T. (2020). The difficulty of statistically analyzing match performance.

4. Muller J. (2021). Introducing the no-touch all-stars. The Athletic. Retrieved from https://theathletic.com/3028824/2021/12/22/introducing-the-no-touch-all-stars/

5. Knutson T. (2017). Revisiting radars. *StatsBomb*. URL: https://statsbomb.com/2017/05/revisiting-radars

6. Koetsier, R. (2021). Percentile Radars/Pizza's. Getting blue fingers. URL: https://www.gettingbluefingers.com/tutorials/RadarPizzaChart

7. CASRAI. (2021). CRediT – Contributor Roles Taxonomy. *CASRAI*. URL: https://casrai.org/credit/

8. Vuong QH. (2020). Reform retractions to make them more transparent. *Nature, 582*(7811), 149.

9. Mastroianni A. (2022). The rise and fall of peer review. *Experimental History*.

10. Seglen PO. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ, 314*(7079), 497.

11. Nguyen CT. (2021). The Seductions of Clarity. *Royal Institute of Philosophy Supplement, 89*, 227–255. doi:10.1017/s1358246121000035

12. Biermann, C. (2019). *Football hackers: The science and art of a data revolution*. Kings Road Publishing.

13. Vuong QH (2017). Open data, open review and open dialogue in making social sciences plausible. *Nature: Scientific Data Updates*.

14. Vuong QH. (2018). The (ir)rational consideration of the cost of science in transition economies. *Nature Human Behaviour, 2*(1), 5.

15. Ho MT, Ho MT, Vuong QH. (2021). Total SciComm: A Strategy for Communicating Open Science. *Publications, 9*(3), 31