

## ***Interactive comment on “Network Analysis of the American Geophysical Union’s Fall Meetings” by Tom Narock et al.***

**M. Parsons (Referee)**

parsons.mark@gmail.com

Received and published: 12 November 2018

### **\*General Comments:\***

This is a clear, well-written, and intriguing paper. It illustrates how LOD and network analysis can be used in creative ways to improve understanding of co-authorship as a proxy for (interdisciplinary) scientific collaboration. It offers some useful insights on what collaboration is currently occurring within the AGU Fall Meeting and how that could be improved. The paper would be more compelling, however, with more nuanced statistical analysis.

Most of the analysis relies on absolute numbers of presentations, components, pairwise comparisons, etc., but these really needed to be considered relatively. There is

C1

little consideration of how the different sizes of the sections, the longevity of authors, and the overall growth of AGU might affect the analyses. For example, the hydrology section probably has the greatest connection with the informatics section simply because it is the largest session, not that it is relatively more engaged with informatics than other sessions. This might also make the Natural Hazards growth even more notable. The authors should also consider the limits of self-classification when discussing keyword use, not to mention how keywords may have changed over time.

The authors also need to decide if they really want to present this as a scientometrics paper or not. On the one hand, they present a clear argument for measuring and assessing data around collaboration in the Earth and space sciences. On the other hand, they are understandably cautious in their assessment because of the limitations of the data. They seem to recognize the limits of scientometrics when they explicitly state that they “are not advocating for any new sort of new metric” (Line 317), but they don’t really explore that. It is worth considering this a “scientometrics” paper, but then the authors must more explicitly grapple with critiques of scientometrics (and infometrics and bibliometrics) and how they are dealing with known shortcomings. The paper as currently written might be considered as an exploration of possible approaches to developing scientometrics. Personally, I would simply avoid the term. I find it fraught and misleading.

It is excellent that the authors provide ready and open access to the data and software used in the study. That said, the data and software should be explicitly cited and listed in the reference list (even if it is self-citation). This is the new norm in Earth and space sciences. See <http://www.copdess.org/enabling-fair-data-project/commitment-to-enabling-fair-data-in-the-earth-space-and-environmental-sciences/>

Overall, I enjoyed reading this paper. I found some useful insights. A bit more careful analysis would make it all the more useful and powerful.

### **\*Specific comments:\***

C2

The title is a bit bland and vague. Maybe borrow from the conclusion and say something like 'Identifying and improving collaborations within AGU using network analysis'.

The abstract reads like an introduction rather than a summary. Please summarize major findings and insights in the abstract and say less about what AGU is and does. Line 74: Good point about ORCIDs. Perhaps AGU could allow members to access their own data and add orcid and resolve conflicts.

Line 140 and table 1: Please indicate when the different sections were established and if there have been any mergers or changes in the sections over the years. Ideally, you need to account for different time periods in making comparisons. The size of the membership and number of presentations in each section must also be considered when making comparisons.

Perhaps I misunderstand, but the following two statements seem to contradict each other. I believe you mean the density is not lower on Line 89 Line 89: "We know that the actual values for network density and connected components are not higher than the values reported here, and they would likely be a bit smaller had we been able to uniquely identify all authors in our dataset."

Line 151: We know that the decline in density for each section is no more than what is shown in Figure 2. Yet, it is likely a bit smaller for each section.

Figure 4 would be more useful if it was shown as ratios or somehow dealt with the huge variance in numbers in each section. Which section is most interconnected relatively?

Section 3.4. How has the keyword hierarchy changed over time? Scenario 3 highlights how this could be significant.

Figure 5: When did NH start? The lack of connection between 2013 and 2016 is as interesting as the growth from 2016 to 2017. The one year jump suggest an event driven cause.

Figures 5-7: Why do they have different time scales? Why not the full 17 years?

C3

Without explanation this suggests cherry picking.

Paragraph starting at Line 246: The authors suggest an uptick in use of the education keyword by PA and U. This does not appear significant in the figure, especially since it's such a short period. Please explain.

Scenario 3 is interesting but speculative. Some discussion of the limits of self-categorization is warranted.

Line 288: Why should single authors be considered worrisome? Bucking a trend is not necessarily a bad thing.

Line 295: "We could make this process more proactive by providing section leadership with connected component data and encouraging connections between specific AGU members. This could range from informal networking events to suggesting session co-conveners." As chair of the Program Committee for Informatics section next year I am interested in how we could do this. It would be nice if you could expand on the idea a bit more. One concern is that we might end up simply reinforcing existing connections rather than expanding the network.

Section 4.1 is interesting. It may be worth noting that AGU has already started clustering sections into "neighborhoods" of related interest. The component analysis could help here, if it could be done in advance of each meeting (Don't want to just follow last years pattern). This is another area where the size of the section would also need to be considered.

Line 311: you need to consider how long the author has been participating in AGU in conjunction with the use of the keyword. Old timers are more likely to have used the keyword repeatedly.

Thank you for the appendix. There appear to be many more interesting scenarios in there. Improved graphical design would make them easier to detect. The coloring of the lines for the different sections in the graphs should be consistent across figures if

C4

possible. It is better to label the actual lines instead of providing a legend or at least list the legend in the same order as the lines appear (top to bottom). Be mindful of the colorblind.

\*Technical corrections:\*

Line 44: delete "only"

Line 13: delete "self-"

Line 383: change ' Yet if' to ', yet it'

---

Interactive comment on Geosci. Commun. Discuss., <https://doi.org/10.5194/gc-2018-15>, 2018.