Geosci. Commun. Discuss.,
https://doi.org/10.5194/gc-2018-15-AC1, 2018

Discussions

Interactive comment on "Network Analysis of the American Geophysical Union's Fall Meetings" *by* Tom Narock et al.

Tom Narock et al.

tnarock@ndm.edu

Received and published: 21 December 2018

Evan,

Thank you for your insightful and helpful review of our paper. We appreciate your time and effort. There are many valuable comments in your review that will enhance our paper. Below, we address each of your points individually.

- Abstract - can you add a line or two about your results and your discussion/conclusion to the abstract?

Yes, we can make this change. This was also suggested by another reviewer and we plan to revise the abstract accordingly.

Printer-friendly version

Discussion paper

- L 24 - can you give a more precise attendance estimate for a year of your analysis or for 2017. I see line 275 of the manuscript - and figure 9 - have a numbers that could be used here.

Our dataset does not contain meeting attendance, only number of presentations given. However, AGU does publish this data online (currently up to 2013) and we can inquire with them for most recent numbers. If we are not able to obtain more recent meeting attendance numbers, we can create a table listing the number of presentations given, which could serve as a proxy for attendance.

- L84-85 - for these 19k cases, did you merge or keep the authors separate?

We merged the 56,155 matches where last name and email address were the same. The additional 19,896 cases where last names matched, initials were a partial match, and email addresses differed (e.g. [T. W. Narock, tom.narock@gsfc.nasa.gov] and [T. Narock, tnarock@ndm.edu] were not merged. The example given here is a known to be a match of the first author. However, we had no means of confirming this for the remaining 19,895 cases and chose to leave them unmerged.

- L90-91 - Couldn't network density go up if a person is duplicated, and therefore a node is actually connected to more nodes (for instance, in figure 1.2, node A and C could be identical people, so the network density would be reported as .67 but actually 1?) also, do some other metrics go up - such as nodes/component?

Yes, you are correct. This is why we show figure 2 as a percent change and list our results as an upper limit. What we show is the "worst case scenario" in which the nodes are not duplicates. The true network density numbers, and other metrics, may in fact be higher than what we show. Unfortunately, given the nature of the dataset this is the best we can show. We think there is still value in showing what the "worst case scenario" looks like. Based on your feedback, we think we should add additional text to clarify this point.

- L109 - i think these are great example of connections that don't appear in the coauthor network diagram - so you can remove 'may' from this sentence.

Thank you. Will do.

- L113-115 - i recommend that the authors make DOIs for their code and data repositories, and cite them in the text using traditional citations (e.g., Narock et al 2018) instead of using links. It seems that the data is already in figshare, so a DOI might already exist.

Great suggestion. We're happy to make this change.

- Line 147 - Shouldn't we expect network density (existing # of edges/possible # of edges) to decrease through time? especially if nodes are added? because for density to remain constant each additional node would need to be added with a (ever larger) number of edges. i.e., each new node adds many new possible edges (the number of new possible edges should equal the number of previously existing nodes), but each node likely only joins the network through a single new edge.

We agree that network density would decrease through time. However, we are surprised by the extent to which it is decreasing (figure 2). If new collaborations were being found at AGU, then we would have expected existing nodes to have new edges at a rate compatible to new nodes being added. For example, in Figure 1.2, nodes A and C have a connection through node B. We anticipated that A and C would eventually connect over the years and this new edge would "counteract" a new node D being added. While the addition of node D would decrease the network density, the edge between A and C would lead to a minimal decrease in network density. Yet, it appears that the new edges are being created at a rate much less than the rate of new nodes joining.

- Table 2 - with so many AGU sections it was difficult for me to keep track of abbreviations and section names. Is it possible for Table 2 to have section names as well?

Yes, we can do this.

- Line 175 - does this mean that your algorithm finds that roughly 30%-50% of AGU presentations are single author? can you randomly check this?

Thank you for catching this. The 50% value does seem high. We will manually verify this.

- Section 3.3 - can you give us some of the info as text here? what are the most connected sections (either sum of connection, or connections normalized by # of nodes), which section has most co-occurences (maybe normalized by section size)? which are the least connected?

Agreed. There is a lot of information buried in the figure. We can revise this section to express these data as text.

- Figure 4: is it possible for you to show this as a shaded matrix, where each sections is listed along the row and column of the matrix, and each cell is color shaded by the number of co-occurences. You would only need to fill in a half of the plot (above or below the diagonal). I think the benefit here would be to visually see that some sections have many connections (i.e., dark shading along a row or column), while others remain unconnected. this is just a suggestion, and may not be feasible/useful.

This is an excellent suggestion. We'd be happy to do this.

- Line 227 - can you determine whether this is a sign of emerging collaborations or the sign of a specific session soliciting abstracts that focus on a specific topic?

This a good question and one we had not considered. We believe there is enough information in the dataset to answer this. We will investigate this further in the revision.

- L274 - is there a way to figure out how many concurrent sessions there are in a given day? might help to contextualize the insanity of the meeting.

Our dataset does have session data and we could compute the average number of

concurrent sessions each day. The average, min, and max could be shown as a second panel on figure 9.

- L286 - i am realizing now that the manuscript presents density change (Fig. 2), but not raw network density numbers for each section (or perhaps i missed it?). that would be interesting to see with regards to this discussion (perhaps in table 2).

We chose not to present raw network density numbers to avoid confusion that they were precise measurements (given the name/email disambiguation issues). Yet, if you think they would be useful we could add them along with additional caveats on node uncertainty.

- L290-299 - these are interesting design considerations. Do you have any concrete examples that you could offer the reader?

Yes, we do. Will include a few examples in the revision.

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