Review of "Exploratory analysis of indicators for open knowledge institutions: a case study of Australian universities"

I think the general topic of this study (open knowledge practices in Australian universities) is an important one. The work that the authors present here is admirable; I suspect that collating all the various indicators must have been very time-consuming. I suspect, in fact, that the dataset itself is the most valuable output from this study. That by any means isn't to dismiss the analyses and conclusions provided in the manuscript, but rather simply reflects the fact that they have produced a large and complex dataset that could plausibly be of great value to future researchers.

In terms of criticisms and potential revisions, my primary suggestion would be that the authors could provide more detail about the rationale for selection and calculation of indicators. After all, nearly every one of the indicators seems to represent a complex piece of information involving many intermediate calculations and decisions. The elaboration of rationale could be at various levels:

1. Providing an explicit definition of "open knowledge" with a rationale for this definition. Currently the authors cite Montgomery et al. (2020), saying that "universities embrace digital OA, but also lead actions in cultivating diversity, equity, transparency and positive changes in society"… This is something a bit like a definition, but the last bit ("positive changes in society") seems like it'd be a nightmare to operationalise. Certainly the authors seem to be adopting a relatively broad definition of open knowledge (e.g., including diversity and equity and so forth); this seems like it could be perfectly defensible, but it would be useful to have an explanation for *why* they think it best to use a broad definition (as opposed to a narrower one that just focuses on things like open access publications, open data, and open source software).
2. Providing more rationale for the selection of indicators. Why, for example, are there indicators for open access publication but not open data sharing or publication of open source software? Why measure diversity of gender but not diversity of age or social class or sexual orientation? In many cases the reason for selecting particular indicators and not others is probably just due to what data is realistically available, which is fine, but it would be useful for the reader to know more about the authors ended up with this selection of indicators.
3. Providing more rationale for more technical decisions made about how to operationalise indicators. For example, why measure the proportion of OA publications rather than the number? Why is it useful to split up the information about proportion of women staff into academic and non-academic roles? Why measure revenue and not surplus/profit? Etc. (I am not suggesting these decisions are "wrong", just that justification for each decision is needed)

This could all take a fair bit of space (some of which might need to go in supplementary materials), but I do think that providing readers with as much information about *how* the indicators were created and *why* they were created in this way would be of great value.

Data analyses

In terms of the analyses that are presented, the authors have clearly considered their choices very carefully and strove to apply sophisticated and rigorous strategies. What I was less confident about was the degree to which the analyses presented answer research questions of real interest in this study.

The justification for the PCA is a notable example here: “This is aimed at providing insight into how information is attributed across the different indicators, and how these indicators relate to a few principle components (PCs)”. This isn’t a very clear rationale, and left me a bit unsure of what the researchers were trying to find out by running a PCA. It’d be good to see instead a specific gap in knowledge identified, and an explanations for why the PCA will fill that gap. Similar goes for the network analysis, and so too the cluster analysis; the analyses are rigorously conducted and reported, but their motivation just wasn’t that clear. Aspects of the descriptive analyses needed justification too: There was a big focus on tests of normality, for example, but it wasn’t really clear why normality of marginal distributions *matters* in the context of this study.

Obviously the analyses are exploratory, which is absolutely fine, but exploratory analyses are still most useful when they’re designed to answer clear, well-motivated research questions. I suspect this could be addressed by either elaborating research questions more clearly, and/or reducing the number of analyses reported.

Specific points

- The first paragraph of the introduction needs citations to back up its points
- “Wherever possible, the data is focussed on the year 2017.” – which indicators are for which years?
- The proportion of outputs with Crossref events seems more a measure of “impact” than openness, no?
- If I’m understanding it correctly, the walk score is primarily determined by how easily pedestrians can access nearby amenities on foot. But the authors seem to be using it as a measure of physical accessibility, which usually signifies things like accessibility *to people with disabilities* (e.g., wheelchair ramps and so forth). More elaboration might be needed here as to why the authors see the walk_score as a measure of accessibility in specific and open knowledge in general. As the results suggest, this variable largely seems to measure *where* universities are located.
- Similarly, there is an accessibility score for the university’s web pages, but from the perspective of staff and students is it not the accessibility of intranet and LMS services that matter more than the external-facing website?
- I’m not sure how much value there is in the ann_rep_div, ann_rep_comm and ann_rep_coord indicators. These don’t seem to measure anything about actual policies or practices, but just how often certain keywords crop up in annual reports. It’s perfectly possible to have an annual report that uses a keyword like “inclusion” many times without enacting any specific policies to meaningfully support inclusion, for example. I’d be concerned, furthermore, that if this set of indicators becomes a target for institutions, that they could easily game the system by inserting keywords into their policies or reports without changing anything substantive.
- Does the network plot in figure 4 display partial or zero-order correlations? Was any regularisation conducted?
- Re. “The generally low correlations between “total_rev” and the OA indicators (except for “oa_bronze”) is an interesting outcome. They are a potential indication that higher revenues at Australian universities do not necessarily translate to higher proportions of OA outputs” – Why would we expect revenue (at an absolute level) to affect the *proportion* of OA outputs? (I can see why we might expect it to influence the absolute *number* of outputs, but that’s a different quantity).

- The authors say that the three-component PCA solution is "is consistent with the OKI evaluation framework suggested by Montgomery et al. (2020).", and then later "These form a proof-of-concept of the OKI evaluation framework proposed by Montgomery et al. (2020)". But the scree plot is actually suggestive of a 2-component solution; the authors also considered a 3-factor solution only to match the 3-factor OKI framework. The findings really can't really be claimed as supportive evidence for the framework.
- Table 1 contains what should be quite crucial information but much of the communication is deferred to citations to external sources (e.g., Kader & Perry, 2007, Huang et al. 2020b, a page on Crossref, etc.). Might it not be preferable to include more information right here? Otherwise the reader needs to do quite a lot of backtracking to get a good idea of how each indicator was actually measured.
- Why does Figure 4 have curved lines? It's a bit hard to read – there are other formats for network plots (e.g., using qgraph in R) that might be more easy for the reader to make sense of.
- Figures 16 and 17 – will it be clear to readers what these dendrograms show? A bit explanation might be needed.

I hope that the authors find these comments useful as they revise this important article. I appreciate the opportunity to provide this review.

PS. Please excuse me not using the preset Peerj reviewing structure, which I don't find to be particularly useful.