Report on the manuscript #30219 by Lopez et al.

1 1. BASIC REPORTING

In this manuscript, the authors examine how the graph edit distance (GED) could be used to evaluate similarity between ecological communities. According to me, this is an interesting topic well described in the introduction, the methods are readable and the findings well reported. Moreover, the manuscript is well structured and conforms to PeerJ standards. Figures are relevant and of good quality. I however I did not understand where the second figure (the pdf file `nyl.pdf`) is supposed to be.

A website to retrieve raw data is mentioned and I was able to use them. However, given what the authors explain lines 111-113, I think it would better if the authors could provide the cleaned data set. Also, as the authors only used one of the data set avaiable on the Interaction Web Database, I recommend to add the following link in the manuscript: https://www.nceas.ucsb.edu/interactionweb/html/thomps_towns.html

As a non-native English speakers, I do not feel comfortable enough to evaluate the quality of the English language used. That said, I think it is overall clear and unambiguous, I was however able to detect one grammatical error in the acknowledgment section, line 274: “We thank to Interaction Web Database of”.

2 2. EXPERIMENTAL DESIGN

2.1 What beta-diversity diversity

It is indeed relevant to quantify how the estimations of dissimilarity based on GED differ from the one based on a common index of similarity (here the Jaccard index). However I believe that the authors should also compare their method to the one developed by Poisot et al. (2012). The authors first mention this study lines 74-75:

> [...], but there is also evidence showing a complete lack of correlation between composition similarity and interaction similarity (Poisot et al., 2012).

First, I think that more is done in this study. T. Poisot and colleagues actually propose a method to quantify the dissimilarity between networks that takes into account the classical beta-diversity as a component of the changes. Second, given that this method actually has a very similar goal to the one the authors develop in this manuscript, to me, they must carry out a meaningful comparison of the two.

2.2 Costs exploration

One important question related to the used of GED is how to determine what are relevant costs in the context of communities dissimilarity? In this manuscript, the authors explore 9 scenarios. I think this is not enough, especially given that AMI (Adjusted Mutual Information metric) is fairly variable among scenarios as highlighted lines 180-181:

> "It should be noted that one of the models where low cost was assigned to interactions (values close to 0), the AMI values were higher (scenario 9; see Fig. 1)."

I believe the authors should further explore how AMI varies with costs scenarios. For instance, they could set 5 scenarios for the cost associated with "Flipping an interaction" (e.g. 0, .25, .5. .75 and 1) and then for each of these scenarios they could create colored matrices that would represent how AMI values vary with a large number of combinations of the two other costs (e.g. from 0 to 1 with an increment of 0.01).

2.3 Details about the method

Even though the methods section is overall clear, I have two questions that require clarifications in the manuscript:

- Why is GED consistently higher than .6?
- Lines 148-149, the authors wrote: "We used 1,000 iterations for the calculations." To what calculations the authors are referring to here and why 1,000 iterations are required?

2.4 3. VALIDITY OF THE FINDINGS

To me, the findings reported here are valid and I have nothing to report in this section. I have however suggested major modifications in the methods section that will likely lead to new findings to report.

2.5 4. General comments

- lines 29-31:

  "The similarity between ecological communities has been one of the central topics in ecology as community structure can impact the functioning, biodiversity, and management of ecosystems."

I believe this sentence should be dropped as it does not add anything to the narrative of the paper.

- line 109: "[..] from ca. The size of these networks varies from ca. 49 to 113 [. . .]" I don't understand why "ca." is being used here.
- line 155: "[. . .] variation in cost operations on AMI [. . .]" what the acronym not l.163
- Errors pertaining to the formatting of references:
  - lines 170-171: "(Core, 2015)" (by the way the R version should be mentioned).

- lines 70-71: “overlapping (i.e. node overlapping index; (Strona and Veech, 2015, Strona et al. 2018; Zhang et al., 2016)”.
- l.138: “(Dehmer, 2010; Ibragim format”

References

Poisot, Timothée, Elsa Canard, David Mouillot, Nicolas Mouquet, and Dominique Gravel. 2012. “The Dissimilarity of Species Interaction Networks.” Edited by Ferenc Jordan. *Ecology Letters* 15 (12): 1353–61. https://doi.org/10.1111/ele.12002.