Dear Editors,

Again, the feedback from editors and reviewers alike is much appreciated. Clearly there were still some lingering issues.

I believe all of that was mentioned by the reviewers below has now been addressed.

Please note that in the proceeding response our text is marked in blue.

Regards,

Matthew DeMaere and Aaron Darling

Editor's Comments

Both reviewers have a relatively minor list of 'fixes' for you to work on. Please note especially Ben's suggestion for a 'Limitations and Future Work' section. I agree that this would be helpful. I judge these to be relatively minor revisions and will not send your paper out for further review, but will evaluate the revised draft and letter addressing the concerns myself when it is ready.

Reviewer 1

Basic reporting

The authors improved the paper significantly, it is much readable now and almost ready for publication.

Experimental design

No Comments

Validity of the findings

No Comments

Comments for the author

Some minor comments:

(1) page 4, line 153: It is still a tautology (since the cardinality of any set is >=0).
Perhaps, it should be ">" or "\neq" instead of ">=".

The paragraph has been revised. However, the inequality (≥) is intentional as it implies that an intersection is “not necessarily empty” rather than it is always non-empty. Clusters within a soft-clustering solution will not always overlap and this needs to be reflected.

(2) page 4, line 159: Mentioning of "1-to-1" is still confusing. It is better to rephrase like "placing a contig into a unique source-genome bin (cluster) is not possible".

Text has been amended to include both the suggested phrase and original term. We feel this will speak to both audience types.

(3) page 4, line 181. It makes sense to define the "harmonic mean".

A reference to the defined term has been added.

(4) page 6, lines 267-268. The meaning of the formula is unclear. Perhaps, it should be something like "(n_i,n_j) is an edge \iff i \neq j".

Amend as suggested.

Reviewer: Ben Woodcroft

Basic reporting

I feel that the manuscript would benefit from some further drafting. Some specific comments:

Fig 4: The HL label in the figure is mangled.

Fixed

Figure 5: The points are stated as “3C-contigs”. I am unclear exactly what is represented in the figure but I suppose the authors mean contig-graphs? If so, the principle components are being derived from a matrix of contig graph vs (sweep parameters from Table 1 where the algorithm choice column is not a factor but given the numeric value of its Fb3)?

Fig 5: It is unclear why some sizes are different

Fig 5: It should be stated that the percentages after the axis labels show the percent of variation explained, if that is what they show.

The three points above together:

The figure legend has been amended with the above comments. The legend neglected to provide the complete term “3C-contig ***graphs***”, which is clearly confusing even to the author.

The double-sized points represent the mean of factored groups (n3c) and it is apparently not possible to disable it. The legend now includes explicit reference to the variation explained.

Figure 6 caption: misspelling of" increased".

Revised

Figure 6. The caption only refers to Fb3, when it should also refer to Pb3 and Rb3.

Revised caption.

143: The first sentence still feels out of place, and the paragraph would flow better if it was moved to after "multiple aspects must be considered".

Thank you. Sentence relocated.

442: There is no d panel in Figure 4.

Revised figure and legend to use (ABCD) tags to agree with text body.

479: Use of the word "parameters" suggests that the runtime is an input of the sweep, it should be replaced/rephrased.

Reworded sentence, though it should be noted that two algorithms (MCL and SR-MCL) did share a "runtime parameter" (inflation) under control in the sweep.

487: The sampling depth is unitless, when it shouldn't be. It might be more directly informative if the number of reads per genome length was reported.

Revised the text. No units are included as we are reporting a count of read-pairs. As such, we have revised the text to make it explicit rather than leave it as an implicit understanding that depth (in the context of a sampling process) connotes "number of samples taken".

We do not feel that normalisation by genome length (i.e. "pairs/bp") would result in a more universal value. Doing so could potentially introduce bias, possibly very particular to the subject genome, and thus complicating the assessment of clustering performance.

This could be explored in future by subjecting many genomes to the described sweep. I have extended "Limitations" to reflect this.

606: It might be clearer to refer to "contig length-weighted" rather than "object-weighted".

Revise the sentence for clarity.

609: I think it can reasonably be concluded that the same conclusions hold for Archaea.

Revised.

Experimental design

No comment.

Validity of the findings

Some further (but brief) enumeration of the assumptions used in the simulation in the “Limitations and Future Work” section would be helpful e.g. the simplicity of the community, error free reads, etc.

Mention of sequencing error, community simplicity and reliance on a single ancestral genome has been added to the “Limitations” sections.

In regards to the software itself, nestly is missing from the list of dependencies (though it is in the pip install list), and I suspect that perl is also needed. Also, the install instructions suggest that “community” is needed, but this python module seems to be an unrelated module, perhaps another was intended. Either way, I was unable to test because no sample.xml file was given and I was unsure of its structure. It would also be helpful to indicate that the “-j N” refers to the number of cores used.

Revised the README.md file for dependencies and other issues. The missing file definition file “config.yaml” was added to master branch. The **community** module was changed to reflect its public name **python-louvain**.