

# Guidelines for reviewing computational papers

Tandy Warnow

July 14, 2019

## 1 Instructions.

Learning to read scholarly papers carefully and critically is a big part of becoming a researcher. This document is designed to assist you in producing a review of a paper, perhaps in the context of a course assignment (e.g., for a course I teach). Since my research is in computational phylogenetics, for the most part the papers you'll be reading and critiquing are either comparing existing methods that are already publicly available or introducing a new method and evaluating it, where the evaluation is usually based on simulations or biological data.

The instructions here emphasize how to critique the paper, and especially how to evaluate the paper's experimental performance study. Approach this assignment as though you were reviewing this paper for a journal. What would you say? Consider the paper in the context of other studies you have read from the same publication period (i.e., the previous years). If the paper was written ten years ago and uses a methodology that was considered reasonable then, don't criticize the authors or the study for not using a better methodology – but do note in your discussion whether the use of a less rigorous methodology impacts the conclusions one can make from the study.

Note also that it isn't possible to do a good critique until you learn how to read scientific papers carefully. So, even if you aren't performing a review of a paper, you can use the questions here to help you learn how to read papers critically: look at the questions here and read the paper so that you can answer the questions. Also, check the supplementary materials if needed to answer the questions. (In fact, when you read a paper, you should *always* read the supplementary materials, since these often have many interesting details that are missing from the main paper.)

## 2 Questions

### 1. What method is described in the paper?

- What is the input to the method, and what is the output?

- Does the method address some specific computational problem (e.g., an optimization problem)? If so, state the problem.
  - Is the method deterministic or does it employ randomness in some fashion?
  - Is the method available as software in some publicly available location? Is that software open source?
  - Are there other methods for the same computational problem? If so, describe them and provide references to the literature where they are presented and/or studied.
2. **Relevance to biological data analysis:** Is there a corresponding use case in biological data analysis? If so, what is that use case? (There may be more than one.) How well does the paper address any of the use cases you consider?
  3. **The method:** To some extent, describe the method. Does it have any guarantees in terms of running time (e.g., it is always polynomial or is it a heuristic search that may need to be stopped at some point)? If it has an optimization problem, does it have guarantees about solving its optimization problem optimally? Does it have any guarantees at all, such as statistical consistency under some evolutionary model? You may also wish to compare the described method to previous methods, if you are aware of any, for the same problem.
  4. **Performance study:** Focus your attention on the evaluation that the paper provides for the method it presents.
    - For the simulation study:
      - How are the data generated? Describe the pipeline for generating the data, to evaluate whether the data generation process makes sense or has flaws. Consider it in the context of other studies you have read from the same publication period (i.e., the previous years). What you are investigating here is mainly two things: are the data biologically realistic, and is there any suggestion that the data were selected to make the method perform well? For example, in simulation studies that are about gene tree estimation or multiple sequence alignment, consider (a) whether a strict molecular clock is imposed, (b) the complexity of the sequence evolution model (e.g., Jukes-Cantor vs. GTR), and (c) dataset size. For species tree estimation from multi-locus datasets, consider the degree of heterogeneity, gene tree estimation error, number of genes, etc. Each of these factors can favor some methods more than others, and some assumptions (e.g., assuming a strict molecular clock or using simple sequence evolution models such as Jukes-Cantor) make the study less realistic. Similarly, studies that are about analysis of sequences generated

by next generation sequencing technologies or third generation sequencing technologies, examine carefully the assumptions of the sequence generation model – these technologies have different error models and results under one technology may not be transferable to another technology.

- What evaluation criteria are used to evaluate methods? Are these standard or atypical? If no standard criterion was used, what is the justification? Note that often a new or non-standard criterion is used, and that can be fine - but it's generally best to also report results under the standard criteria.
  - What alternative methods were evaluated? If not, why not? Was any justification given? If yes, which methods were used, and were these the best-in-class at the time of publication? Were the alternative methods run to optimize performance, or run in some way that degrades their performance in a significant way?
  - Is statistical significance evaluated?
  - What does the simulation study show? And how does this match what the author claims? Are the differences big enough to matter?
- For biological data:
    - Given that it is much harder to evaluate accuracy on biological data, it can be difficult to comment on this part of the paper. Say whatever you think is reasonable, but mainly focus on the simulation study.

#### 5. Awareness of the literature at the date of publication?

- Are there other methods that were published before this time that the author should have known about and discussed, and which are relevant to the study?
- Are there other published approaches for evaluating the same kind of method published before this time that the author should have known about and discussed, and which are relevant to the study?

#### 6. Other questions

- What does the author claim to establish, and what do you think of this claim?
- Are the data and software used in the study easily findable? Publicly available? (Note that if the author says the data are available upon request, that's not really the same as easily findable and publicly available.)
- To what extent is the study reproducible? (For example, are the specific commands and software version numbers provided?) How difficult and time-consuming would it be to reproduce the study?

- What are the strongest points for the paper?
- What are the weakest points for the paper?

### 3 Final comments

Now that you have read and carefully evaluated the paper, think about what you learned and how it influences your thinking about further research, as well as about writing styles.

1. Did you like the paper? Why or why not?
2. What aspects of the writing were good, and which ones were not so good?
3. If you are asked to make specific recommendations to the author for changes, what suggestions would you make?
4. Would you recommend the paper for publication in a journal or conference? If so, which ones? If not, why not?
5. How would you design a study to evaluate the method that the author provides? What would the properties of the data be that you would use to evaluate the method? What other methods would you compare it to? What criteria would you use? Your design of the study should be based on the biological use case you think is most important. Since computational issues are likely to be an issue, describe what you would do if you had full access to some heavy duty supercomputer (e.g., a few months of uninterrupted time on Blue Waters) and then what you would do if you only had something less powerful (e.g., several hundred hours on the Campus Cluster).
6. What other papers have you read that address the same (or very similar) question and biological use-case, for which the method presented in the paper would be relevant. Are those other studies better or worse, and why?