Dear Dr Kundinger,

Thank you for submitting your paper "Efficient and Scalable Bipartite Matching through Fast Beta Linkage (fabl)" for possible publication in Bayesian Analysis.
It has now been carefully reviewed and my decision is: Major revision required.

These are the comments I received from the Editor who has handled your paper:

The paper has been reviewed by a referee and an associate editor and I have read the manuscript independently, before looking at the reports. When I confront my notes and overall impression with the reviewers' comments, I find myself in substantial agreement with the reviewers.

As indicated by both reviewers, the work expands on previous work by Sadinle, introducing computational advances that are achieved by relaxing some requirements in the original formulation of the problem. The close connections with the previous work and the extent of the novel contributions are not adequately explained. This makes it hard for the reader to understand what is new and why the new contributions are valuable. A careful editing of the presentation is needed to address this shortcoming.

Regarding the presentation, the AE points out a number of problems with notation, typos, and various inconsistencies. Independently, I found many similar problems. The intersection between what I found and what the AE found is non-empty and so is the symmetric difference. Also, the AE found some issues that I did not find and vice versa. So, I am almost certain that an additional review would uncover more problems. Now, the results presented in the paper seem plausible as do the broad strokes of the derivations, but I must admit that I was not able to follow all the details. This, in large part, was because of the issues that I just mentioned, and I believe that most readers would find it difficult to follow the developments, as these issues are exceedingly distracting.

There are problems with the notation, which in places is not defined, in others is used before being defined, and in others yet is used inconsistently. For example, I_obs() is never defined, n_12(Z) first appears on p. 5, but is not defined (in passing) until the bottom of p. 7, and it seems to become D on p. 8. Throughout the manuscript, n_1 and n_2 are used interchangeably with n_A and n_B, sometimes in the same section, as it happens, for example, in Section 3.1. The loss function on p. 9 is out of Sadinle's paper and it inherits the typo therein (\theta_11, if Z_j, \hat{Z_j}, ...). In the statement of Lemma 1 on p.13, B comes out of nowhere. (Thinking I had missed something, I went back to the previous pages only to find out that B would be then defined in the proof.) Formatting of all displayed math must be checked and fixed, especially as far as punctuation is concerned, as commas seem to be missing in multiple places. For example, in Equation (1), there should be a comma after 0 and after 1, and this is by no means an isolated

occurrence.

Confusion can also arise from lack of clarity in the exposition. Take comment 4 from the AE, for example. Whose marriage certificates and whose birth records are involved? What records are in X_1 and what records are in X_2? The reader should not be expected to go back to Newcombe's paper to clarify the issue.

This is not an exhaustive list, but only a set of examples meant to point out what makes the paper hard to read and how it can be improved. Many more specific points are raised in the AE's report and here are a couple of other minor issues I noticed:

p. 12, second paragraph: insert "is" after Section 3.1

p. 15: the first sentence of the last full paragraph is either missing a verb ("we observe"?) or it includes an extra "that."

As I said, I do not believe that we, as reviewers, were able to uncover all the problems, and I strongly encourage the authors to do their part, as they should, to improve the presentation and eliminate all typos and inconsistencies.

The AE finds the simulation studies to be incomplete and gives detailed suggestions on how those should be improved. The AE also gives important suggestions about other aspects of the manuscript that must be carefully considered and addressed. All of the AE's comments are right on the mark.

To summarize, the paper contains an interesting algorithmic contribution that can speed up calculations at the expense of relaxing some of the modeling conditions, without much adverse impact on the resulting inferences. This aspect should be emphasized in a revised presentation. The presentation should also make clearer the close connections to the relevant work by Sadinle, and all the issues that I mentioned above should be ironed out.

The authors should prepare a careful and substantive revision that remains within the editorial limit of 25 pages and that answers the various comments satisfactorily. The authors should upload the revised manuscript together with a document detailing how they addressed the reviewers' comments.

==========

You can find the reviewers reports by logging in at EJMS. To submit your revision, please log in to EJMS and submit it
as a revised file to original submission. Please also include a detailed description of how you

addressed all the points raised by the reviewers.

Thank you for considering Bayesian Analysis as a venue for your work.

Sincerely,
Mark Steel
Editor-in-Chief, Bayesian Analysis