AE Comments

1 AE Comments with responses

The comments to the AE need to be extremely precise and accurate. I have attempted to write one in painful detail to provide guidance and an example.

1. Line 111. Is this missing at random or something else? This needs to be clarified (AE Comment 1). This needs to be addressed with extreme care regarding the precise wording.
   This distinction between truly missing at random and missing completely at random does not matter due to two assumptions. Following Sadinle (2017), we assume that the fields in the comparison vectors are conditionally independent given the coreference matrix and we assume that the missing comparisons are ignorable (Sadinle (2017). Therefore, our likelihood is the same as in Sadinle (2017) (see equations (3) and (4) of his paper). As stated by Sadinle (2017), "we work under the assumption of ignorability of the missingness mechanism for the comparisons so that we can base our inferences on the marginal distribution of the observed comparisons (Little and Rubin, p. 90)." We refer to Sadinle (2017), Section 4.2, page 604 for further details.

2. Appendix 8.2: I appreciate the streamlining of the discussion of the proposed algorithms and think it was a good choice to move the full derivation to an appendix. However, I do not follow the re-expression of the pmf for $\Gamma_{\cdot j}$. First, the opening square brackets are still misplaced throughout. In the second line of this derivation, the authors divide the expression by a product of the element-wise conditional probabilities of a match in the comparison vector (u), raised by an indicator that the element of the comparison vector equals a particular value. Perhaps I misunderstand, but I believe this product is not constant in $\Gamma_{\cdot j}$ or u, and so the total expression is not proportional to the line above, as interpreted as a pmf. I think that the end effect is that the last line of the newly expressed pmf is missing a factor of
   Is the new proof, Brian, missing an indicator over the cluster assignment? I believe that it should match those of Sadinle (2017), Aleshin and Guendel (2022), and Wortman (2019). Where is the indicator regarding the cluster assignment that you condition on? Specifically, the update for each gamma will depend upon your cluster assignment, so I believe there is a

typo for this reason. (To help with notation, Jody calls this S_k, Sadinle calls it Z_{-j}, and Serge calls it $\mathcal{C}_{-j}$). Could you carefully check this please? I think this question is related to the next point/comment by the AE.

3. Appendix 8.2: Thank you for including the details on integrating out pi from the full conditionals. They surprised me. I had assumed that the authors had integrated out pi in the prior for Z, as this is what Sadinle (2017) had done to form the "beta prior for bipartite matchings". Does the alternate approach presented in this paper provide a different algorithm than directly integrating the prior distribution? Is the presented alternative approach justified?
Thank you for noting your surprise regarding our proposal. We do not extend the approach of Sadinle (2017), but relax the one-to-one assumption for computational reasons. This motivates the choice of our prior (instead of the beta prior for bipartite matchings). Note: I do not believe that we should be integrating out the prior distribution. I think we did this at the confusion from the first revision of the AE. We should instead just update the conditionals to form a Gibbs update, which will make the algorithm justified. If you believe that the algorithm is correct, we will need to convince the AE why it is a valid Markov chain or rather Gibbs update of the partition. It will not be justified unless the full conditionals of the Gibbs sampler contain a sequential update or one proves that this is a justified Gibbs sampler. This was the point that Serge mentioned in his email. Currently, it is not a valid Gibbs sampler. I attempt to explain this more clearly below. Looking at Pitman, Pruenster, and the BNP literature, I cannot find a reference regarding why the Markov chain would be valid in this case. If you were to prove this, it would be a new result (to my knowledge).

   It would need to be something like this, which comes from Aleshin-Guendel and Steorts (2023), which extends the general clustering literature. The partition is equivalent to the co-reference matrix or linkage structure. The partition approach is more widely understood by the Bayesian nonparametrics community (or statistics community), which is why we are using this terminology. (They will not easily understand the linkage structure or coreference matrix without connections being made for them).

   > " The most common way to update the partition, $\mathcal{C}$, in Bayesian entity resolution is the Gibbs update [**?** **?** **?** **?** **?** **?** **?**], which originated in the Bayesian clustering literature [see e.g. **?**]. We emphasize that this is only referring how the *partition* is being updated.
   >
   > The Gibbs update works by sampling the cluster assignment for each record $i \in [R]$ sequentially. To describe the update we will represent a partition $\mathcal{C}$ by a vector of labels $\boldsymbol{Z} = (Z_1, \ldots, Z_R)$, where $Z_i \in \{1, \ldots, R\}$ such that $Z_i = Z_j$ if records i and j are in the same cluster in $\mathcal{C}$. Suppose that we have already sampled

the cluster assignment for the first $i-1$ records. Sampling the cluster assignment for record i proceeds as follows:

(a) Let $\boldsymbol{Z}_{-i}^{[t]} = (Z_1^{[t]}, \ldots, Z_{i-1}^{[t]}, Z_{i+1}^{[t-1]}, \ldots, Z_R^{[t-1]})$ denote the current partition of $[R]$ after removing record i.

(b) Let $\boldsymbol{c}$ denote the union of the unique labels occupied in $\boldsymbol{Z}_{-i}^{[t]}$ and one unoccupied label. Each of these labels correspond to a cluster to which record i could be assigned, with the unoccupied cluster representing the creation of new cluster with only record i.

(c) Sample record i's cluster assignment, by sampling the cluster $c \in \boldsymbol{c}$ with probability proportional to $p(Z_i^{[t]} = c \mid f(\boldsymbol{X}), \boldsymbol{Z}_{-i}^{[t]}, \boldsymbol{\theta}^{[t]})$.

"

4. Sections 4.1 and 4.2 (I have gone through this, however, I suggest that we all go through this many times and write out responses very clearly). I am doing my best to get through this as quickly as possible and make recommendations.