

[PAKDD2017](#)**21st Pacific Asia Conference on Knowledge Discovery and Data Mining 2017**

May 23–26, 2017, Jeju Island, Korea

Reviews For Paper**Paper ID** 107**Title** Privacy-Preserving High-Quality Record Linkage using Random Hashing**Masked Reviewer ID:** Assigned_Reviewer_1**Review:****Question**

Overall Rating	Reject
Reviewer Expertise	Expert
Detailed Comments	<p>In this paper, the authors presented a new PPRL approach based on the random hashing function. The authors also provide the analytics of the results on different databases. Some comments and suggestions are given below.</p> <ol style="list-style-type: none"> 1. This paper is similar with the another submitted article called "Efficient Cryptanalysis of Bloom Filters for Privacy-Preserving Record Linkage". I am wondering the differences between those two papers. 2. The contents should be well-organized since I could not find the major contributions and the designed approach (should be written in pseudo-code or presented it in more specific way) 3. If this paper has better performance, than the another submitted paper becomes useless? 4. The contributions of this paper can only show that the authors only evaluate the algorithms in two hashing methods? (random hash and BFs). I think the contribution is too weak. 5. The frequency-based way is not suitable in real-world cases since many factors are involved in the databases.

Masked Reviewer ID: Assigned_Reviewer_2**Review:****Question**

Overall Rating	Accept
Reviewer Expertise	Knowledgeable
Detailed Comments	<p>This paper present very clearly a privacy-preserving scheme, argues solidly for its quality, complexity and performance against state of the art BF methods, and finally complements it with a very thorough experimental section. The paper description is detailed and enables the reproducibility of the results, not only because of the transparency of the exposition but because of the availability of the code. Only a few errors appear in the paper which are minor typos.</p> <p>On Page 2, "of string values [16] However, a" is missing a dot and</p>

	<p>should be "of string values [16]. However, a"</p> <p>There could be some issues on the complexity, as in the method several steps are quadratic "and compares all pairs in each block ". And clearly, reducing some parameters would have a high impact on the quality, both, on the quality of the privacy-protection and second on the quality of the matching.</p> <p>I am not so convinced that the statement "Assuming $m \ll N$, the cost of sending the classification vectors dominates all other communication costs.", this statement should have been emphasised in the experimental section.</p>
--	--

Masked Reviewer ID: Assigned_Reviewer_3

Review:

Question

Overall Rating	Weak Reject
Reviewer Expertise	Knowledgeable
Detailed Comments	<p>This paper presented a method to do record linkage using random hashing. The proposed method build on the top of random maximum margin hashing, and utilized a number of SVM models to achieve improved linkage quality. The proposed method is technically sound, but my main concerns are with the privacy level of the proposed method. How the privacy can be evaluated? Does it satisfy K-anonymity or differential privacy? The experiment utilises the frequencies to evaluate the privacy, but how this is related was not elaborated adequately.</p>

[PAKDD2017](#)**21st Pacific Asia Conference on Knowledge Discovery and Data Mining 2017**

May 23–26, 2017, Jeju Island, Korea

Meta-Reviews For Paper**Paper ID** 107**Title** Privacy-Preserving High-Quality Record Linkage using Random Hashing**Masked Meta-Reviewer ID:** Meta_Reviewer_1**Meta-Reviews:****Question**

Detailed Comments	In this paper, the authors propose a privacy-preserving record linkage (PPRL) approach based on random hashing. The approach can improve privacy characteristics and the linkage quality compared to commonly used Bloom lter based PPRL techniques. The analysis of the privacy characteristics of the proposed approach is provided, and an experimental evaluation on several real-world data sets is given to demonstrate the validity of the proposed method. The representation of the paper needs to improve further, since some sentences are difficult to understand.
-------------------	--

Masked Meta-Reviewer ID: Meta_Reviewer_2**Meta-Reviews:****Question**

Detailed Comments	Message from the PC co-chairs: If a paper received two negative reviews, we decided to reject the paper, even though the meta-reviewer suggested acceptance, in order to avoid the bias among the meta-reviewers.
-------------------	---