

SUPPLEMENTARY DOCUMENT

This document contains two figures that have been omitted from our main submission to EDBT 2024.

- (1) Distribution of expansion ratios when using a lower threshold of 0.7 as Jaccard similarity between the columns in quadruples. This is shown in Figure 1. As can be seen, the expansion ratios are still very high and the figure is very similar to Figure 8 in the original paper that uses a threshold of 0.9.
- (2) The breakdown of useful vs accidental pairs of tables according to the sizes of key columns. This is shown in Table 1. There is no clear correlation between the table sizes and the frequencies of the usability vs accidental pairs.

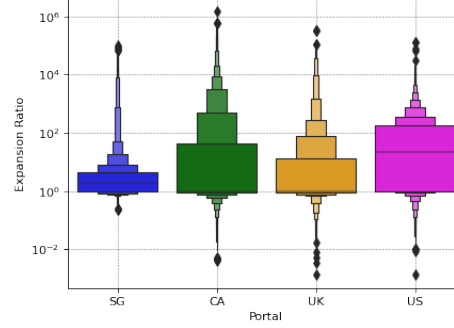


Figure 1: Distribution of expansion ratios of joinable pairs with threshold ≥ 0.7 .

Portal	key column	Join Result			
		U-Acc	R-Acc	total	useful
CA	$10 \leq t < 100$	21.57%	60.78%	82.35%	17.65%
	$100 \leq t < 1000$	35.29%	50.98%	86.27%	13.73%
	$t \geq 1000$	50.98%	39.22%	90.20%	9.80%
UK	$10 \leq t < 100$	25.49%	54.91%	80.40%	19.60%
	$100 \leq t < 1000$	33.33%	50.98%	84.31%	15.69%
	$t \geq 1000$	36.73%	40.82%	77.55%	22.45%
US	$10 \leq t < 100$	73.47%	14.29%	87.76%	12.24%
	$100 \leq t < 1000$	72.00%	22.00%	94.00%	6.00%
	$t \geq 1000$	43.14%	35.29%	78.43%	21.57%

Table 1: Distribution of accidental vs useful labels with regards to different table size groups.