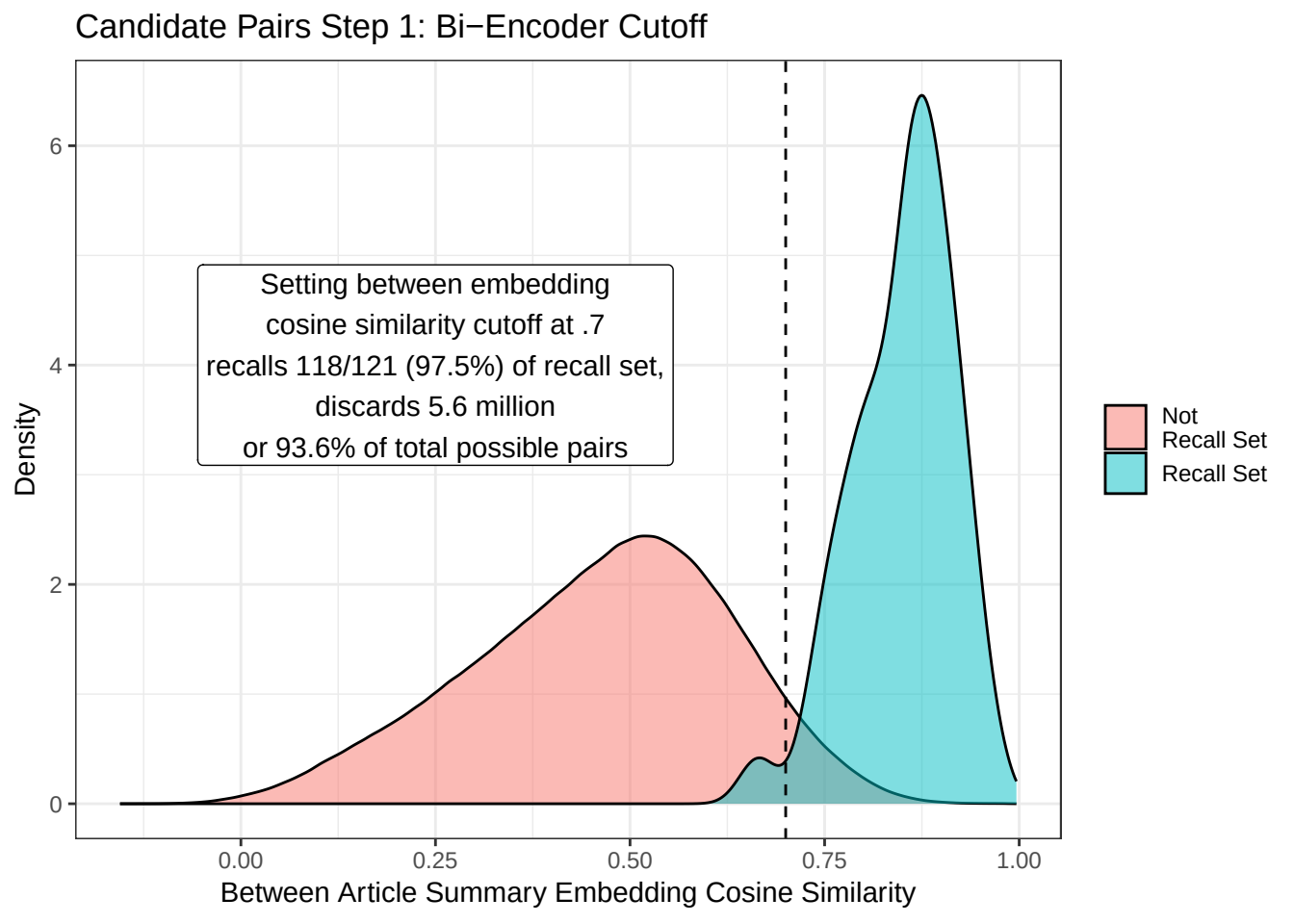
Candidate Pairs Step 1: Bi−Encoder Cutoff
Setting between embedding
cosine similarity cutoff at .7
recalls 118/121 (97.5%) of recall set,
discards 5.6 million
or 93.6% of total possible pairs
Density
0
2
4
6
0.00
0.25
0.50
0.75
1.00
Between Article Summary Embedding Cosine Similarity
Not
Recall Set
Recall Set