Capturing the confidence of ϕ_{author} [Longformer] when predicting the author centric labels on \mathcal{D}_{test} 700 Author-centric spectrum 600 k = Papers without artifacts (n = 109) 500 1.0 k = Papers with artifacts 400 that aren't permanantly archived (n = 553) 300 k =Papers with artifacts 200 that are permanantly archived (n = 136) 100 Fraction of positives $\mathbb{I}(y_{test,\,i}=k)$ Hypothetical ϕ_{author} with perfect calibration 8.0 0.0 0.2 0.4 0.6 8.0 1.0 Papers without artifacts 300 0.6 250 200 150 100 50 0.0 0.2 0.4 0.6 8.0 1.0 Papers with artifacts that aren't permanantly archived 350 300 250 200 150 0.0 100 50 0.0 0.0 0.2 0.6 0.4 8.0 1.0 0.4 0.6 8.0 1.0 Papers with artifacts Mean predicted probability \hat{p}_k for ϕ_{author} that are permanantly archived