I. From-scratch Name Disambiguation (SND)

Evaluation Metric --- Macro Pairwise-F1:

$$\text{PairwisePrecision} = \frac{\#PairsCorrectlyPredictedToSameAuthor}{\#TotalPairsPredictedToSameAuthor}$$

$$\text{PairwiseRecall} = \frac{\#PairsCorrectlyPredictedToSameAuthor}{\#TotalPairsToSameAuthor}$$

$$\text{PairwiseF1} = \frac{2 \times PairwisePrecision \times PairwiseRecall}{PairwisePrecision + PairwiseRecall}$$

Evaluation sample:
Assume that an author α has 12 papers, and predicted model finally clusters the 12 papers into two clusters ---A, B: A has 8 papers while B has 4 papers, we can get the predicted model performance as following:

$$\#PairsCorrectlyPredictedToSameAuthor = \frac{8 \times 7}{2} + \frac{4 \times 3}{2} = 34$$

$$\#TotalPairsPredictedToSameAuthor = \frac{8 \times 7}{2} + \frac{4 \times 3}{2} = 34$$

$$\#TotalPairsToSameAuthor = \frac{12 \times 11}{2} = 66$$

$$\text{PairwisePrecision} = \frac{34}{34} \times 100\% = 100.00\%$$

$$\text{PairwiseRecall} = \frac{34}{66} \times 100\% = 51.52\%$$

II. Real-time Name Disambiguation (RND or CND)

Evaluation Metric --- WeightedF1:

For each author:

$$\text{Precision} = \frac{\#CorrectlyPredictedToTheAuthor}{\#TotalPredictedToTheAuthor}$$

$$\text{Recall} = \frac{\#CorrectlyPredictedToTheAuthor}{\#UnassignedPaperOfTheAuthor}$$

$$\text{Weight} = \frac{\#UnassignedPaperOfTheAuthor}{\#TotalUnassignedPaper}$$

For all authors (M is the number of authors):

$$\text{WeightedPrecision} = \sum_{i=1}^{M} Precision_i \times weight_i$$

$$\text{WeightedRecall} = \sum_{i=1}^{M} Recall_i \times weight_i$$

$$\text{WeightedF1} = \frac{2 \times \text{WeightedPrecision} \times WeightedRecall}{WeightedPrecision + WeightedRecall}$$

Evaluation Sample:
Assume that there are total 100 new(unassigned) papers, 12 of which belong to author α. Predicted model finally assigns 10 of the 100 papers to the author α, 8 of them are correct. Then, we can get the predicted model performance as following:

$\#CorrectlyPredictedToTheAuthor = 8$

$\#TotalPredictedToTheAuthor = 10$

$\#UnassignedPaperOfTheAuthor = 12$

$\#TotalUnassignedPaper = 100$

$Precision = \frac{8}{10} \times 100\% = 80.00\%$

$Recall = \frac{8}{12} \times 100\% = 66.67\%$

$Weight = \frac{12}{100} = 0.12$

III. Incorrect Assignment Detection (IND)

Evaluation Metric --- We adopt Area Under ROC Curve (**AUC**), broadly adopted in anomaly detection, and Mean Average Precision (**MAP**), which pays more attention to the rankings of the anomalies, as the evaluation metrics.

For each author:

$$\text{Weight} = \frac{\#ErrorsOfTheAuthor}{\#TotalErrors}$$

For all authors (M is the number of authors):

$$\text{WeightedAUC} = \sum_{i=1}^{M} AUC_i \times weight_i$$

$$\text{WeightedMAP} = \sum_{i=1}^{M} MAP_i \times weight_i$$