(a) E-SHARC

First Speaker Assignment

Test Recording

Diarization Output

Embeddings Extraction

ETDNN Model

X_t

Similarity Matrix

$\left(H_t^0 = \begin{bmatrix} X_t \\ X_t \end{bmatrix}, S_t^0\right)$

Graph Initialization

(G_t^0)

(G_t^{m+1})

(H_t^{m+1}, S_t^{m+1})

Graph Generation

$m \leftarrow m + 1$

Aggregation

GNN Module

Scoring $(\hat{p}_{ij}^m)$

Clustering (C_t^m)

repeat till convergence

(b) E-SHARC-Overlap

Second Speaker Assignment

Final Diarization Output

First Spk.

Second Spk.

cluster1

cluster2

cluster3

overlap

clean

Output Graph

Scoring

GNN Module

k-nn graph (only inter-cluster edges)

External Overlap Detection

(a) MAE of recordings with ">10" spks count

(b) MAE of recordings with "5-10" spks count

(c) MAE of recordings with "1-4" spks count