

Cell Linking Benchmark

The main objective of this document is to describe in detail the task to be solved and the evaluation protocol to be followed for all submissions to the Cell Linking Benchmark.

Task To Be Solved

Given an incomplete set of segmentation masks (**ERR SEG**), an algorithm is expected to establish tracklets over those masks and complete possible temporal gaps between the established tracklets to form lineage trees (**ALG RES**). Note that **ERR SEG** always covers more than 99% of objects available in the reference tracking annotation (**TRA GT**) and does not contain any spurious segmentation masks for real sequences, and it is complete for all computer-generated sequences. Because it is strictly forbidden to completely segment the whole image domain to detect missing objects in **ERR SEG**, the reference tracks with no vertex in **ERR SEG** and the prolongation of the reference tracklets before/after their first/last known vertex in **ERR SEG** are excluded from the task assigned as well as from evaluation. See Figure 1 for more details.

Evaluation Protocol To Be Followed

The linking accuracy measure (**LNK**) assesses how accurately each given object has been followed in successive frames by comparing the acyclic oriented graph computed by an evaluated algorithm with the reference graph given by the reference tracking annotation. To this end, **LNK** takes **TRA GT**, **ERR SEG**, and **ALG RES** as its input, creates two intermediate acyclic oriented graphs (**Pruned TRA GT** and **Modified ALG RES**) with synchronized sets of vertices and zero penalty assigned to the vertex set synchronization, and calculates a normalized [Acyclic Oriented Graph Matching measure for association](#) (AOGM-A) over the pair of intermediate graphs. Numerically, **LNK** is defined as

$$\mathbf{LNK} = 1 - \min(\text{AOGM-A}, \text{AOGM-A}_0) / \text{AOGM-A}_0$$

where AOGM-A_0 reflects the cost of creating **Pruned TRA GT** from its vertices only (i.e., it is the number of links in **Pruned TRA GT** weighted by the cost for adding an edge). The minimum operator in the numerator prevents the **LNK** score from being negative in the case when it is cheaper to create **Pruned TRA GT** from its vertices only than to transform **Modified ALG RES** into **Pruned TRA GT**. The normalization ensures that **LNK** always falls in the $[0,1]$ interval, with higher values corresponding to better linking performance.

The biological accuracy measure (**BIO**) averages biologically inspired measures applicable to a given dataset, which are evaluated over the pair of intermediate graphs:

- Complete Tracks (**CT**) measures the fraction of reference cell tracks that a given algorithm can reconstruct entirely from the frame in which they appear to the frame in which they disappear. It is especially relevant when a perfect reconstruction of the cell lineages is required.
- Track Fractions (**TF**) averages, for all detected tracks, the fraction of the longest continuously matching algorithm-generated tracklet with respect to the reference track. Intuitively, this can be interpreted as the fraction of an average cell's trajectory that an algorithm reconstructs correctly once the cell has been detected.
- Branching correctness (**BC(*i*)**) measures the efficiency of a given algorithm at detecting division events with a tolerance of *i* frames.

- Cell cycle accuracy (**CCA**) measures how accurate an algorithm is at correctly reconstructing the length of cell cycles (i.e., the time between two consecutive divisions).

These biologically inspired measures, and thus **BIO** too, take values in the $[0,1]$ interval, with higher values corresponding to better biological performance.

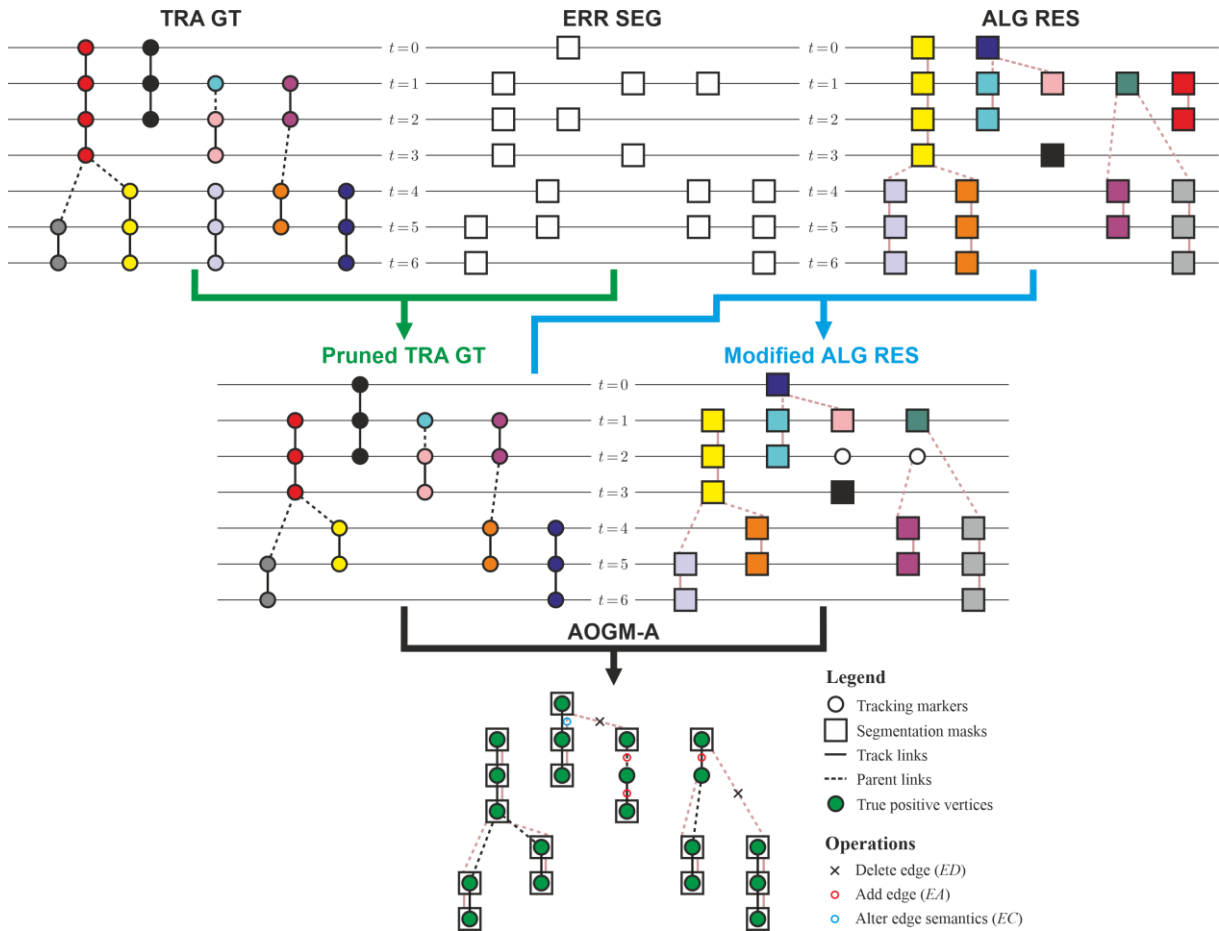


Fig. 1. Visualization of the task to be solved and the evaluation protocol to be followed within the Cell Linking Benchmark. **TRA GT**, **ERR SEG**, and **ALG RES** refer to the reference tracking annotation, the input set of segmentation masks, and the output of a competing algorithm for a particular sequence. The pruning of **TRA GT** consists in removing complete tracks without any corresponding vertex in **ERR SEG** (e.g., the metal blue track from $t=4$ to $t=6$ in **TRA GT**) and in removing the prolongation of tracklets before/after their first/last known vertex in **ERR SEG** (e.g., the red vertex at $t=0$ or the yellow vertex at $t=6$ in **TRA GT**). The modification of **ALG RES** aims at synchronizing the sets of vertices in **Pruned TRA GT** and **ALG RES**, thus allowing straightforward evaluation of the Acyclic Oriented Graph Matching measure for association (**AOGM-A**). This modification consists in removing all spurious vertices and their incident edges from **ALG RES** (e.g., the yellow vertex at $t=0$ in **ALG RES**) and adding missing vertices to it (e.g., the pink and magenta vertices at $t=2$ in **Pruned TRA GT**) with inherent creation of consistent parental links (e.g., the parental link from the yellow vertex at $t=3$ to the metal blue vertex at $t=5$ or from the white vertex at $t=2$ to the magenta vertex at $t=4$ in **Modified ALG RES**).