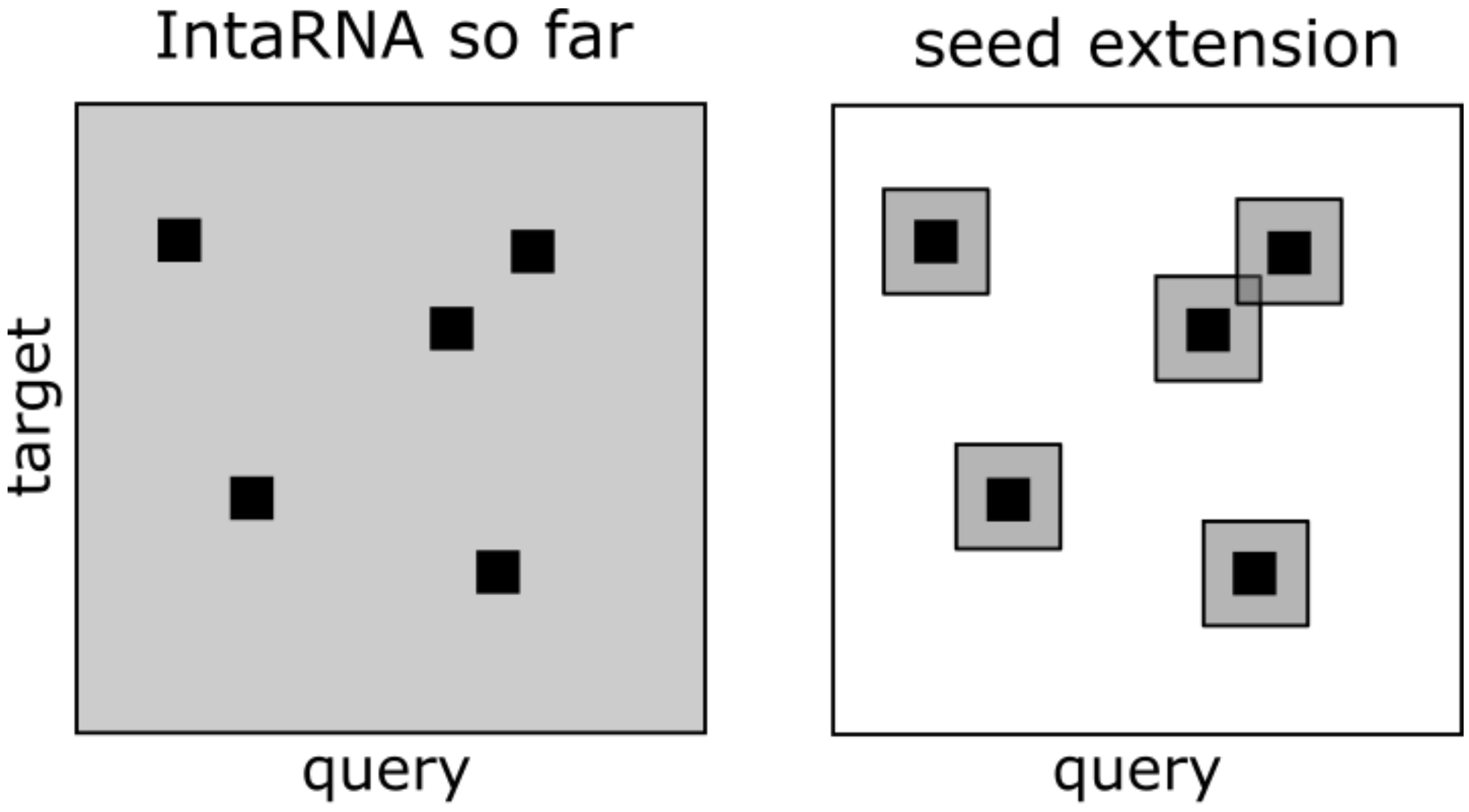
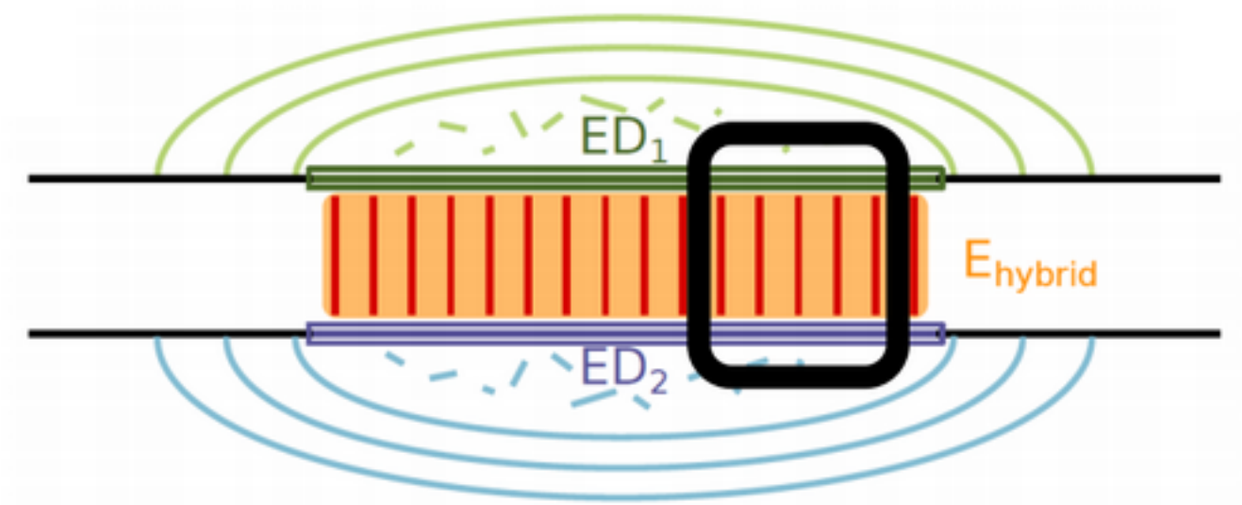
RNA-RNA interaction prediction using seed extension

Frank Gelhausen

Motivation


IntaRNA so far
seed extension
target
query
query
Seed
eg: sTarPicker
RiBlast
RiSearch2

Exact memory efficient method



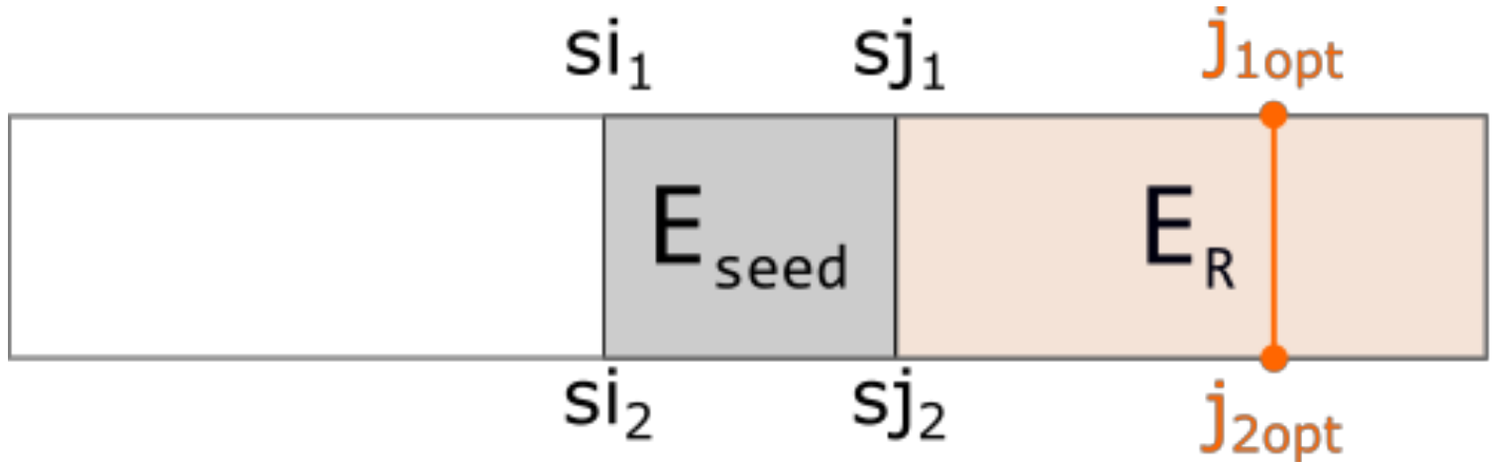
Search minimum over all seeds using:

$$\min \left[E_{1oop} \; E_L \right] + \left[E_{seed} \right] + \min \left[E_R \; E_{1oop} \right]$$

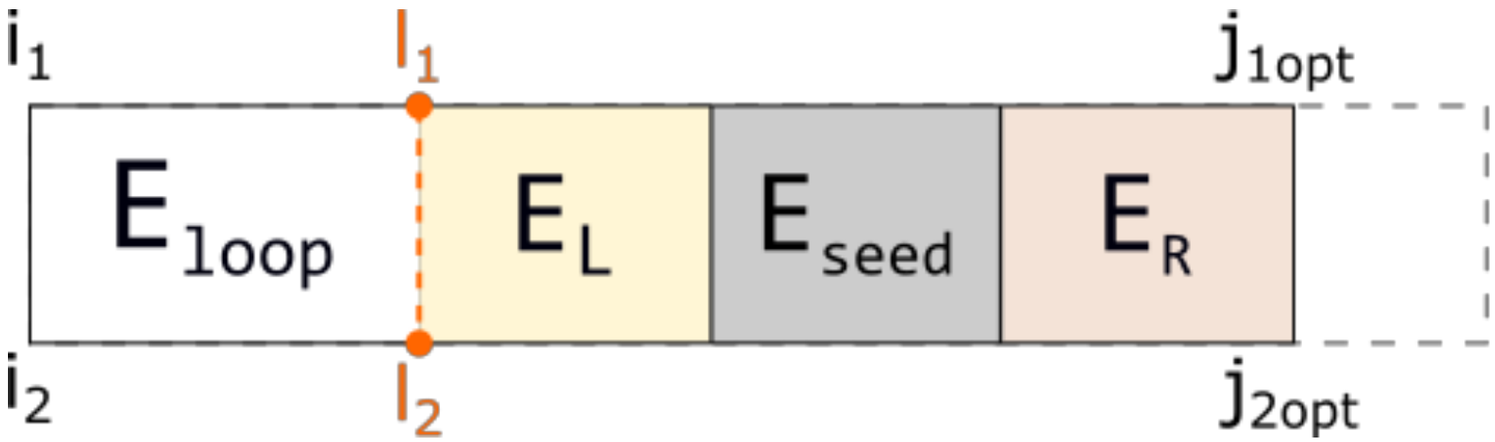
i_1 i_2 l_1 l_2 r_1 r_2 j_1 j_2

Heuristic method

- First find j_1 and j_2 that minimize E_R

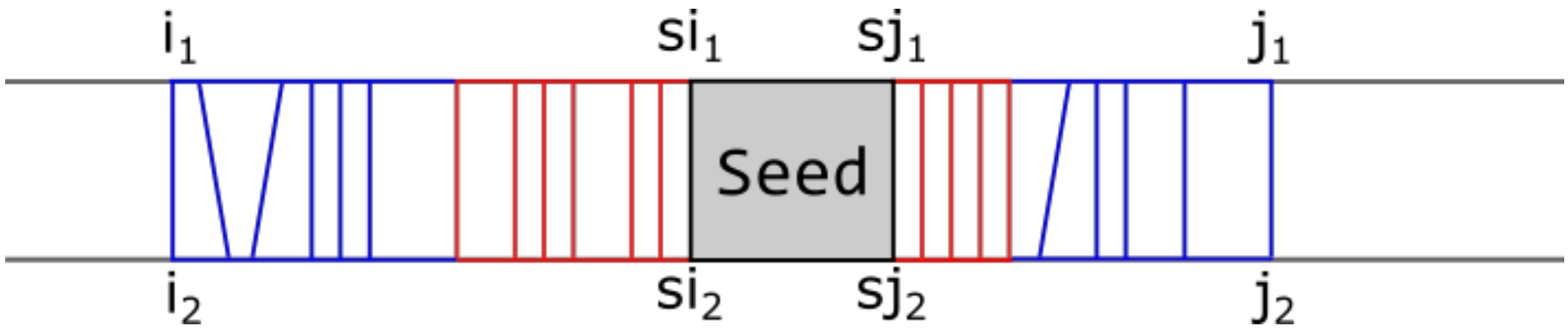


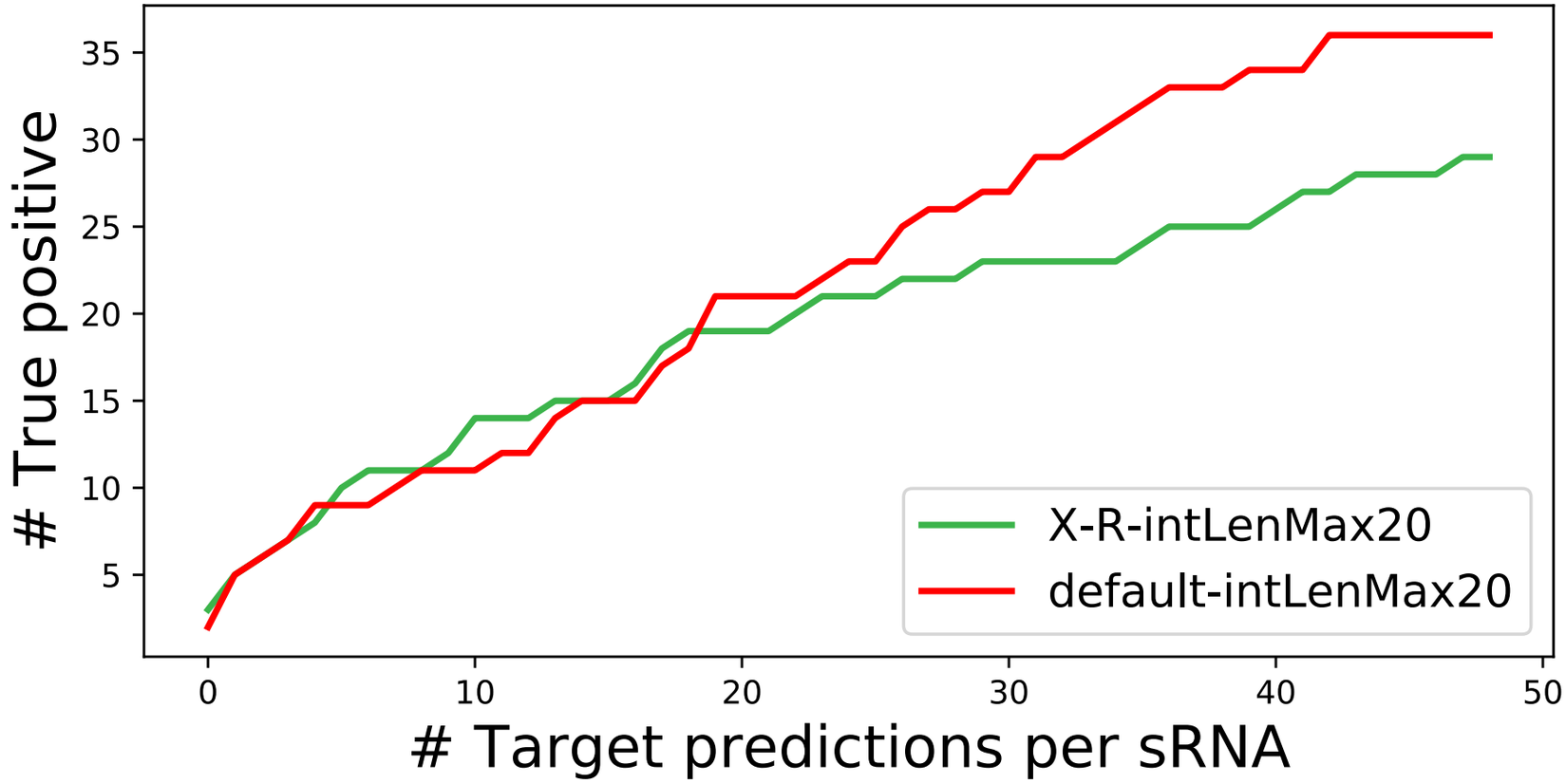
- Then minimize over entire interaction up to j_{1opt}, j_{2opt}



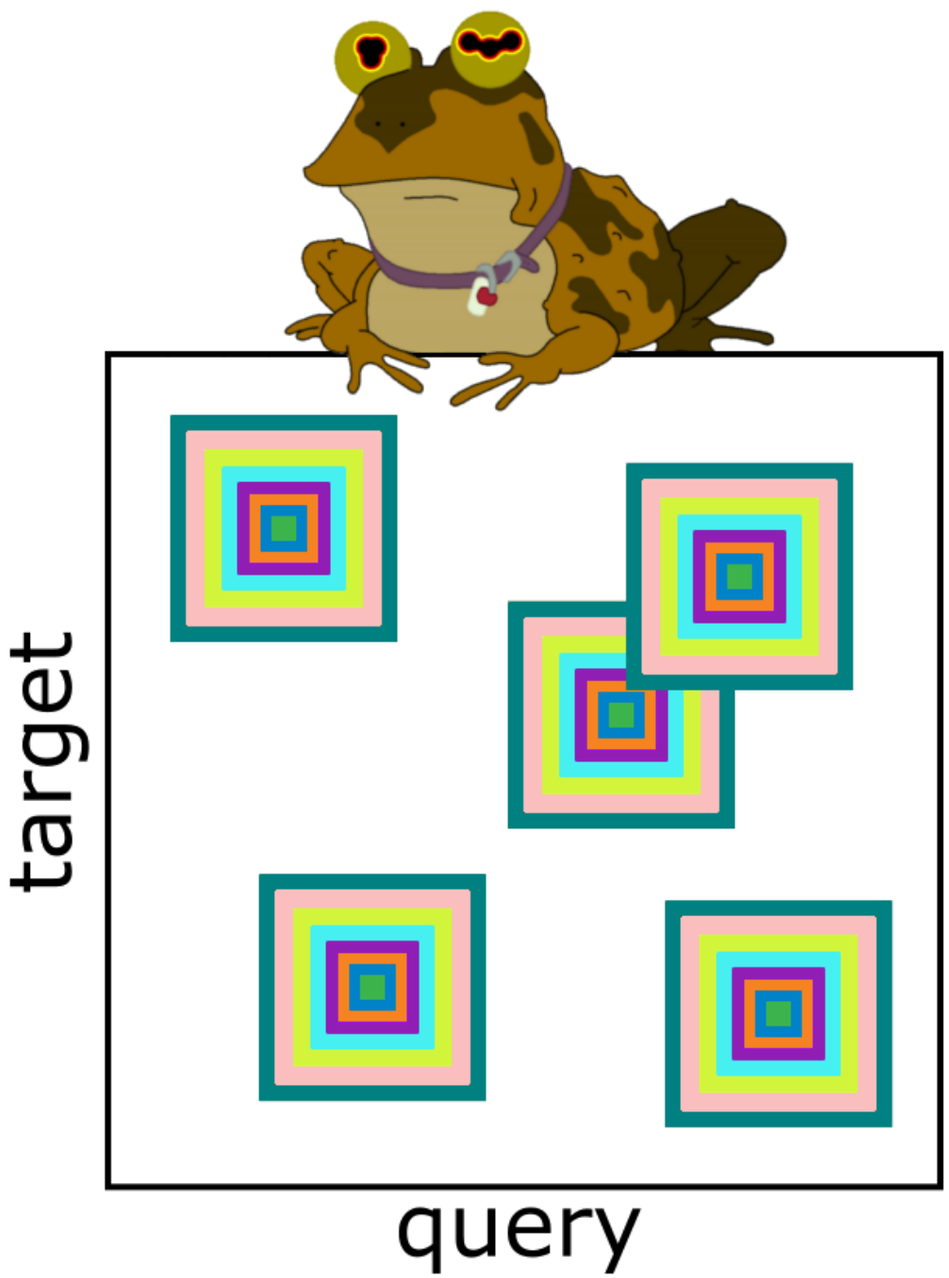
RiBlast method

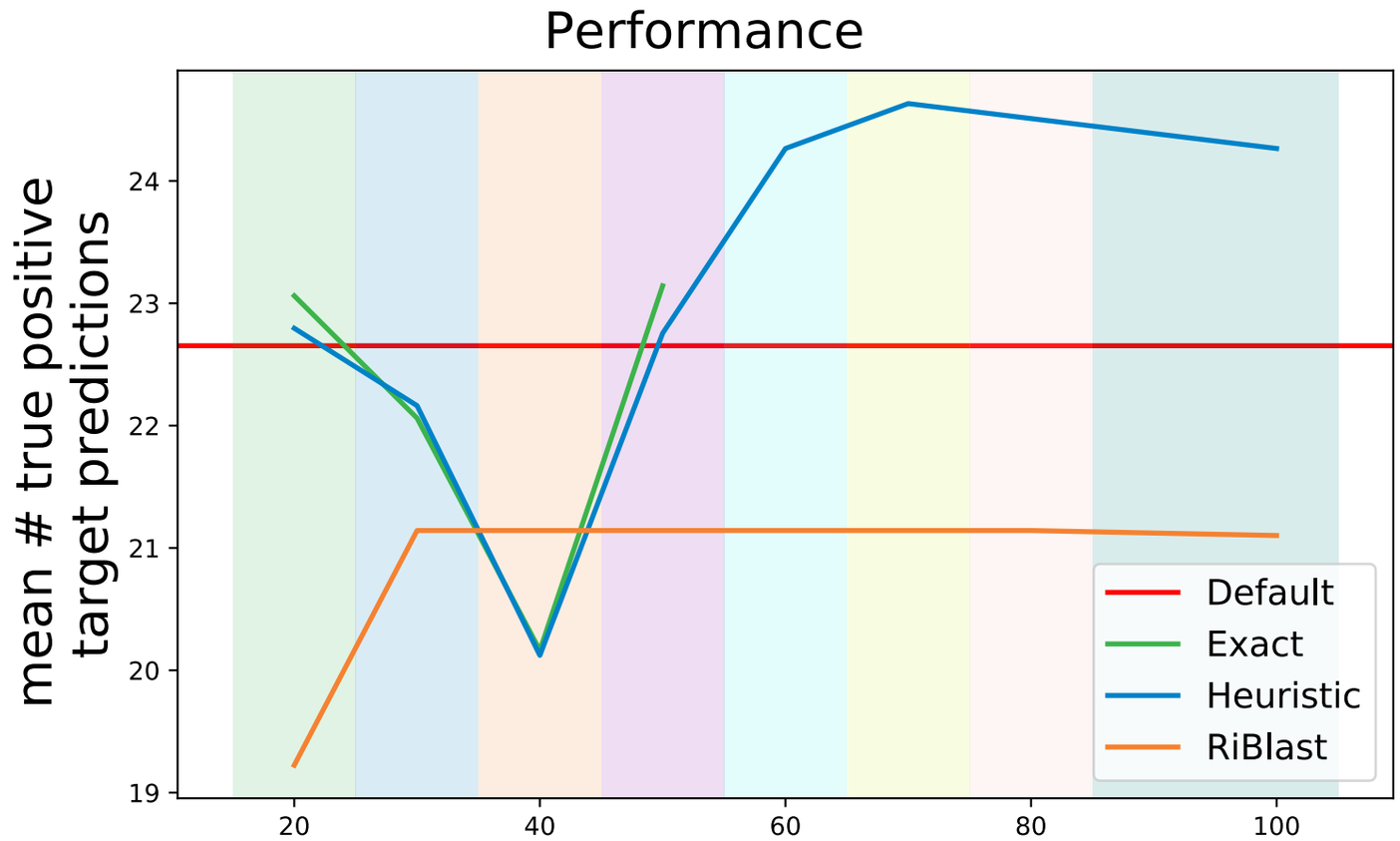
First do a parallel extension, then thorough extension

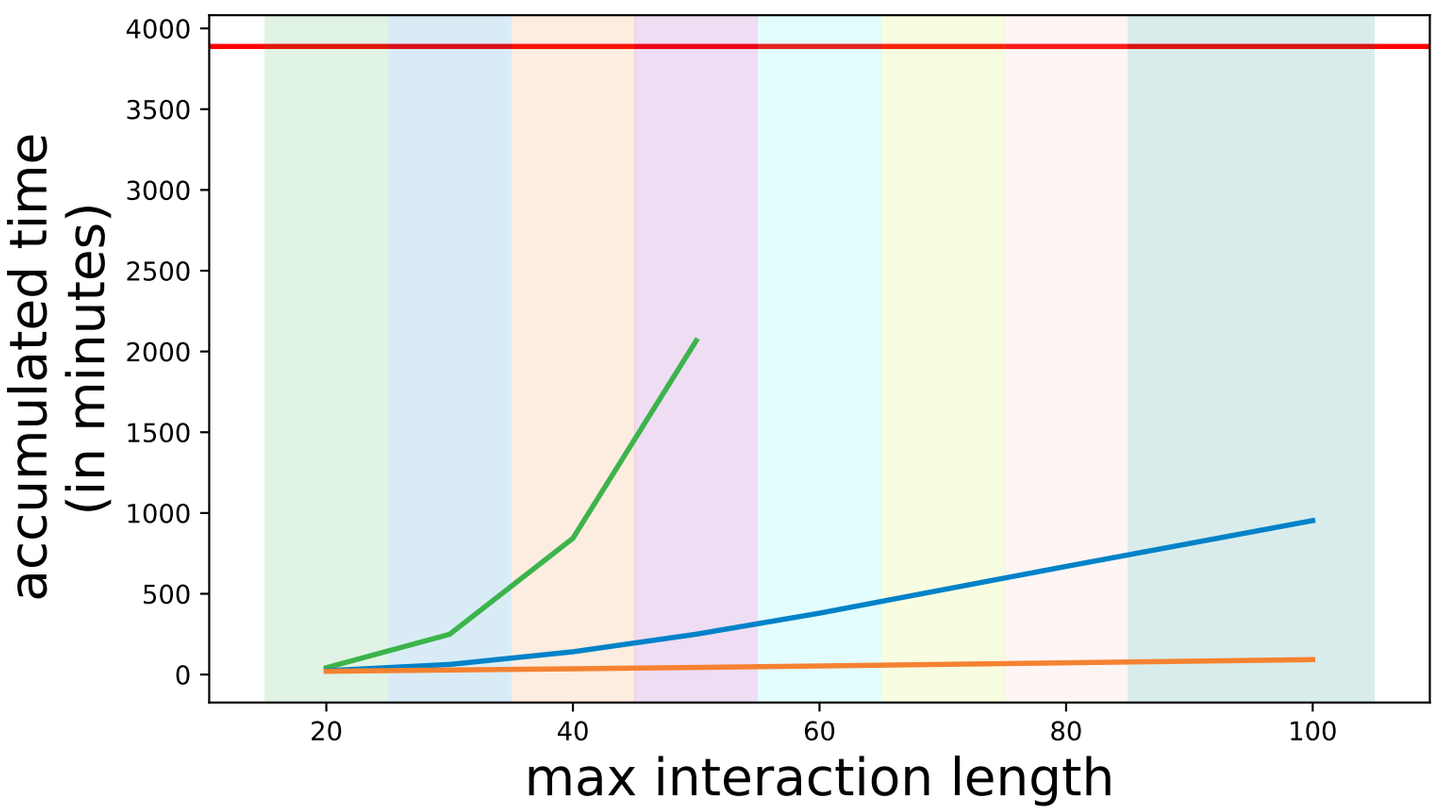




Length-dependence







Partition function

Goal: Find probabilities of interactions

$$P_{i,j} = \frac{Z_{i,j}}{Z}$$

$$Z_{i,j} = \sum_{S_A} \left(\begin{array}{|c|c|c|c|} \hline \text{no seed} & S_A & S' & S'' \\ \hline \end{array} \right)$$

i j

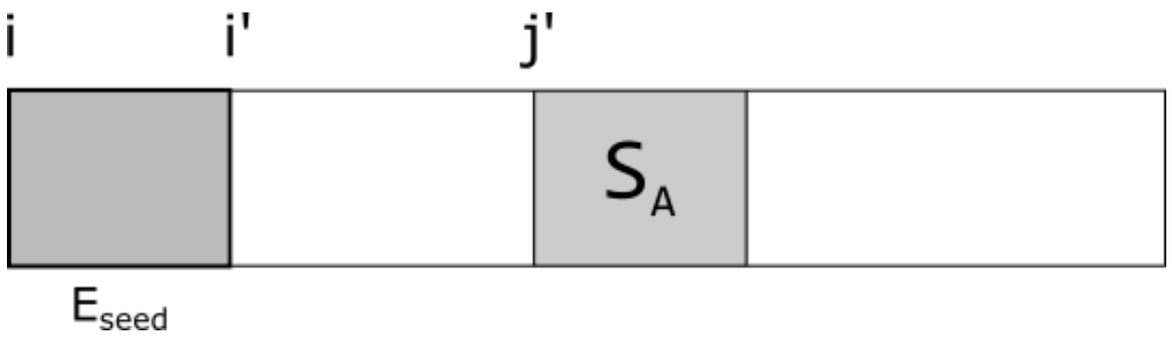
$$Z = \sum Z_{i,j}$$

$$Z_{i,j} = \sum_{S_A} \Big(\text{i} \quad \boxed{\text{no seed}} \; \boxed{S_A} \; \boxed{\text{S'}} \; \boxed{\text{S''}} \quad \text{j} \Big)$$

Idea: remove seed for i

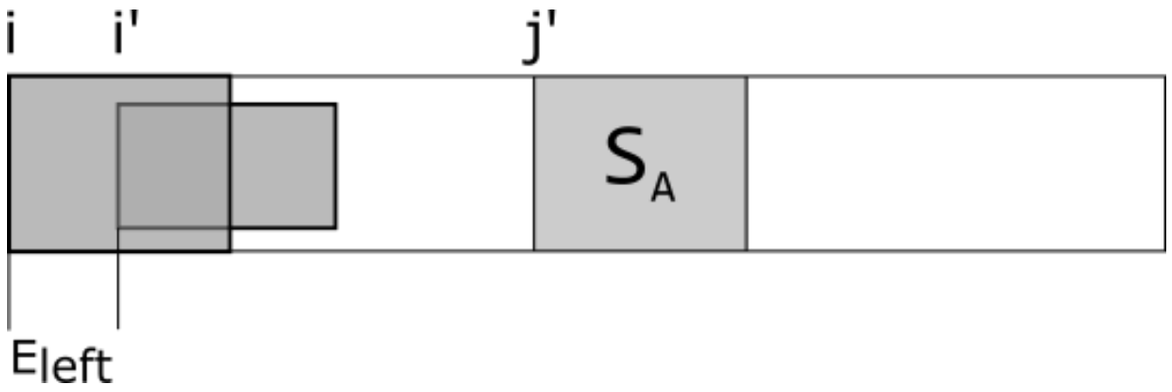
Cases:

A)



$$Z_{i,j'} - w(E_{seed}) \cdot Z_{i',j'}$$

B)



$$Z_{i,j'} - w(E_{left}) \cdot Z_{i',j'}$$