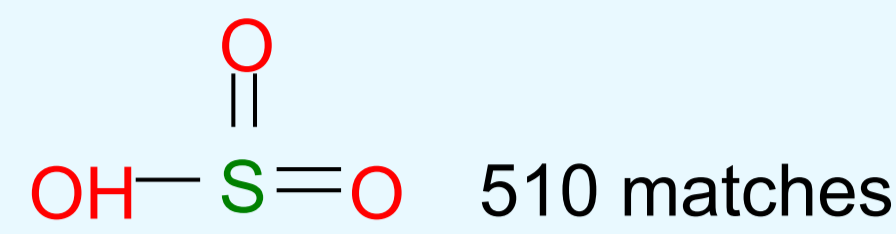


# Graph Query Reformulation with Diversity

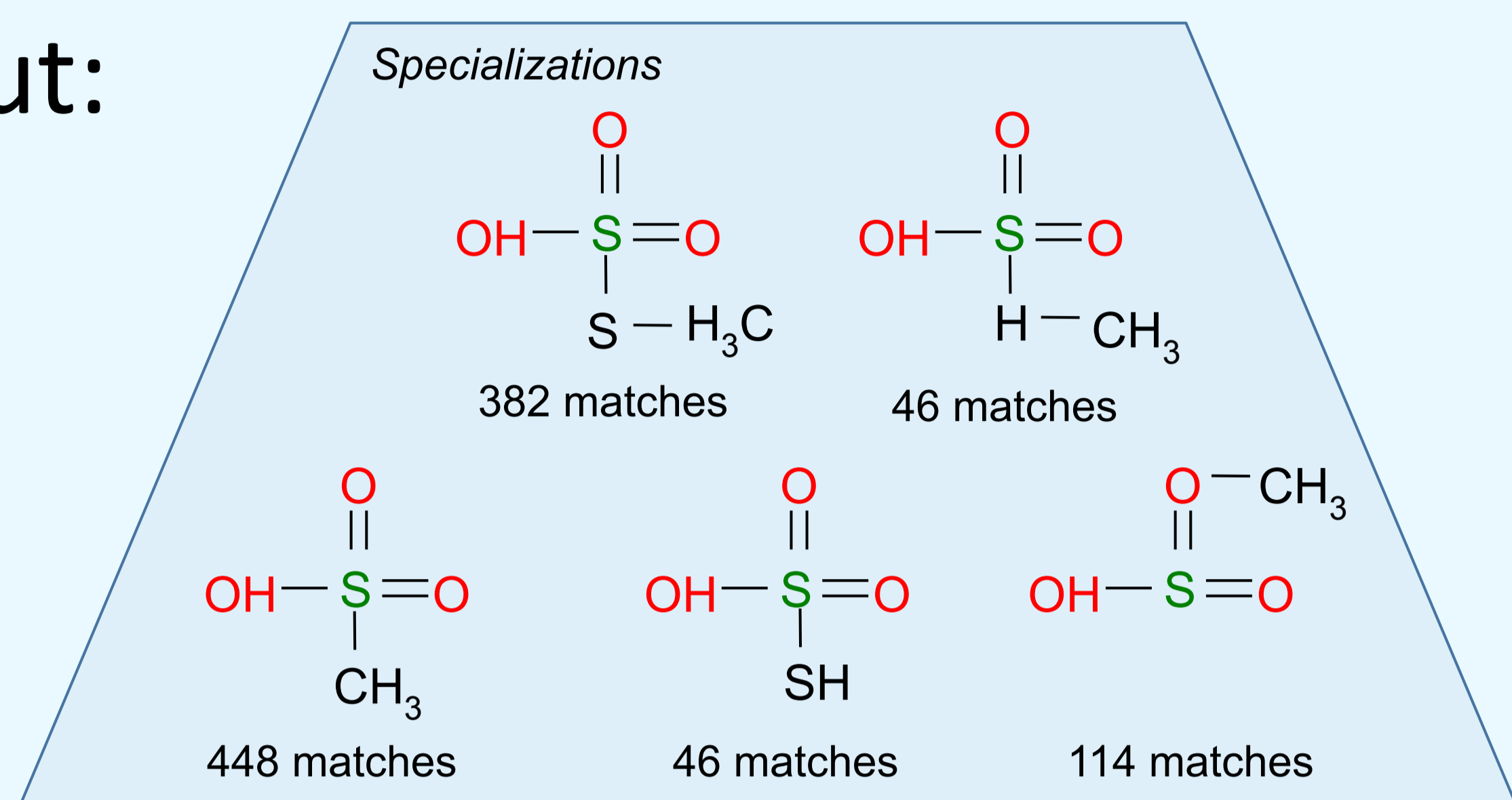
Davide Mottin (U of Trento), Francesco Bonchi (Yahoo Labs, Barcelona), Francesco Gullo (?)

## Motivation

Input:



Output:



- The user is overwhelmed by the number of results
- Wants a compact representation the structure in the support set
- Does not have a precise idea of the structure of interest

## Problem

Find **k meaningful** reformulations that:

1. Span **all** the results

$$cov(Q) = \left| \bigcup_{Q' \in Q} D_{Q'} \right|$$

2. Present **different** aspects of the results

$$div(Q', Q'') = |D_{Q'} \cup D_{Q''}| - |D_{Q'} \cap D_{Q''}|$$

## Objective Function

$$f(Q) = cov(Q) + \lambda \sum_{Q', Q'' \in Q} div(Q', Q'')$$

$$Q^* = \arg \max_{Q \subseteq \mathcal{S}_Q} f(Q)$$

subject to  $|Q| = k.$

## Greedy Algorithm

**While**  $k$  reformulations are not found

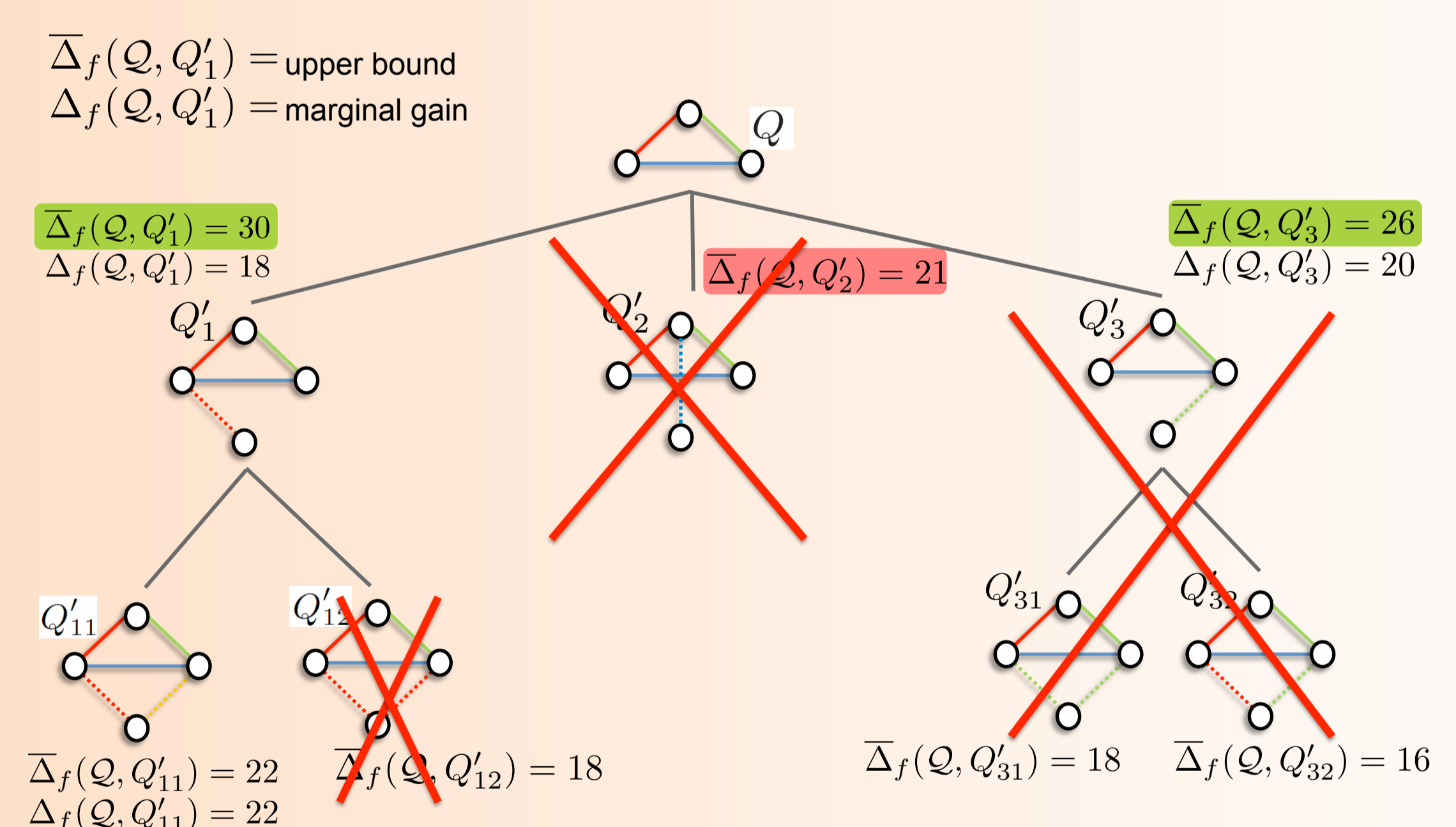
1. **Find** the reformulation leading to the maximum increment of the objective function (*marginal potential gain*)

**[#P-complete step]**

2. **Add** the reformulation to the results

## Fast\_MMPG

**Idea:** Prune the search space computing an upper bound on the marginal potential gain



## Performance

Compare Fast\_MMPG with the top-k most frequent reformulations in terms of time and quality

Query C=O

