Conservative atomizing

Input: A directed graph $G = (V, A)$ with articles V, citations A, a subset $W \subseteq V$ of articles in an author's profile merged according to a partition $\mathcal{P}$ of W, and numbers $h, k \in \mathbb{N}$.

Find: A partition $\mathcal{R}$ of W such that
i) $\forall R \in \mathcal{R} : |R| = 1 \vee R \in \mathcal{P}$,
ii) $|\mathcal{P} \setminus \mathcal{R}| \leq k$,
iii) the h-index of $\mathcal{R}$ is at least h.

How to count citations of merged articles $R \in \mathcal{R}$? **See right column**.

Theoretical results

(among others)

Conservative atomizing
i) is linear-time solvable for sumCite and unionCite,
ii) W[1]-hard parameterized by $h + k + s$ for fusionCite,
where $s = \max_{P \in \mathcal{P}} |P|$.

Conservative atomizing is not solvable in time $f(h + k + s) \cdot \text{poly}(|V|)$ for any function f, under ETH.

If Google **used fusionCite**, manipulation would be **harder**!

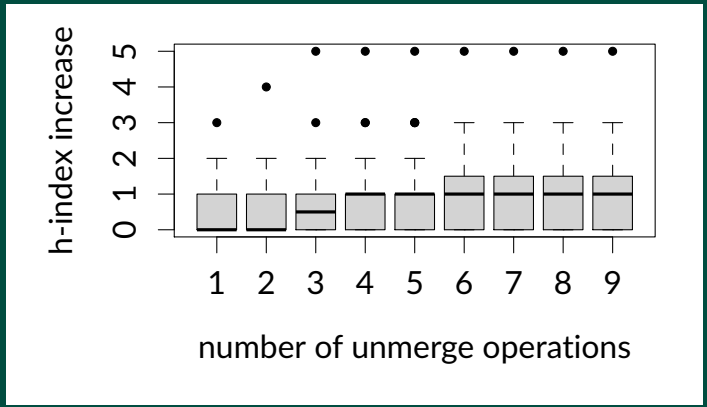
Experiments

On Google Scholar profiles of early-career IJCAI 2013 authors and "AI's 10 to watch" 2011 and 2013.

H-index manipulation by undoing merges

R. van Bevern, C. Komusiewicz, H. Molter, R. Niedermeier, M. Sorge, T. Walsh
Quantitative Science Studies 1(4):1529–1552, 2020.

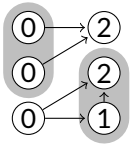
Optimally unmerging articles to maximize the h-index in Google Scholar is easy.



A more intuitive citation measure for merged articles would make it harder.

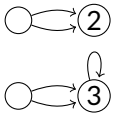
Citation measures

Individual citation counts **before merging the shaded articles**.

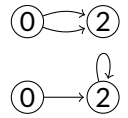


After merging the shaded articles:

sumCite. Citation count is the sum of individual citation counts:



unionCite. Used by Google Scholar. Citation count is the number of citing individual articles:



fusionCite. More intuitive. Treat each merged article as one, i. e., remove double- and self-citations: