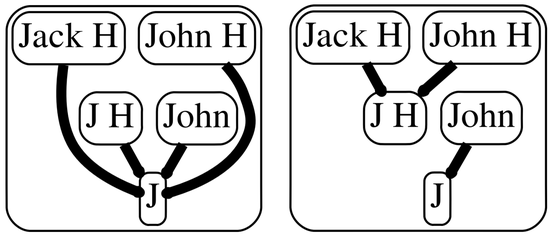
Lattice-based Progressive Author Disambiguation

Tobias Backes & Stefan Dietze

1 Author **names are rountinely used for blocking** in *Author Disambiguation:*

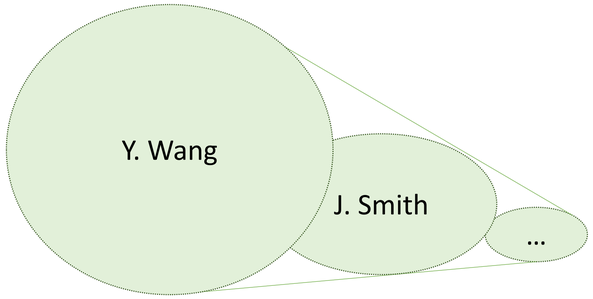


first initial — all initials

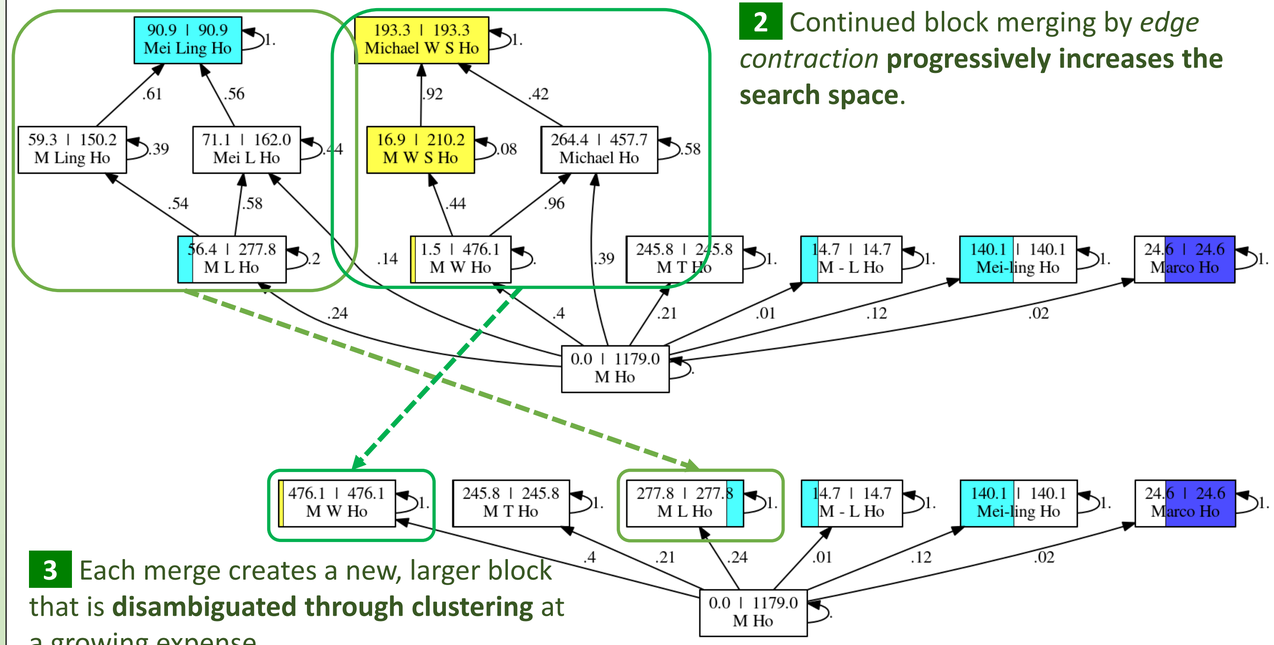
2 But **which names are equivalent**?

John Doe = Jack Doe ?
J. H. Doe ≠ J. Doe ?

3 Also, how to avoid **oversize blocks**?



1 The *subset partial order* arranges author names in a semi-lattice to encode **name-based relationships**.



2 Continued block merging by *edge contraction* **progressively increases the search space**.

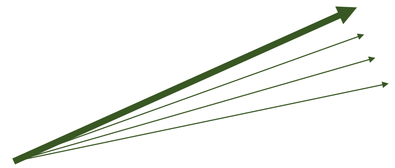
3 Each merge creates a new, larger block that is **disambiguated through clustering** at a growing expense.

4 Different *edge weighting* schemes lead to **different blocking progressions**.

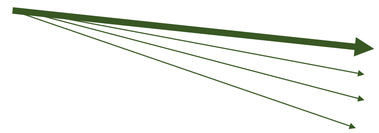
progression: growing blocks
increasing recall, decreasing precision

1 First progressive blocking method specifically designed for author disambiguation.

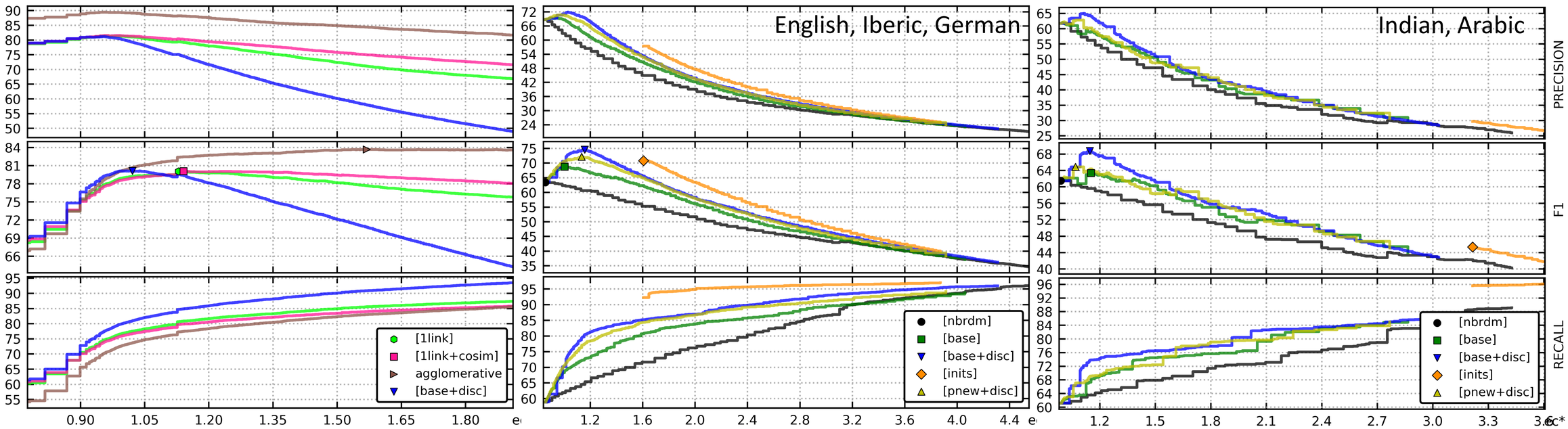
2 Outperforms conventional static approaches as well as simpler progressive baselines.



3 Fastest increase in Recall with growing block sizes.



4 Slowest drop in Precision.



gesis Leibniz-Institut für Sozialwissenschaften

hhu Heinrich Heine Universität Düsseldorf