

# Most similar maximal clique query on large graphs

Yun PENG, Yitong XU, Huawei ZHAO, Zhizheng ZHOU

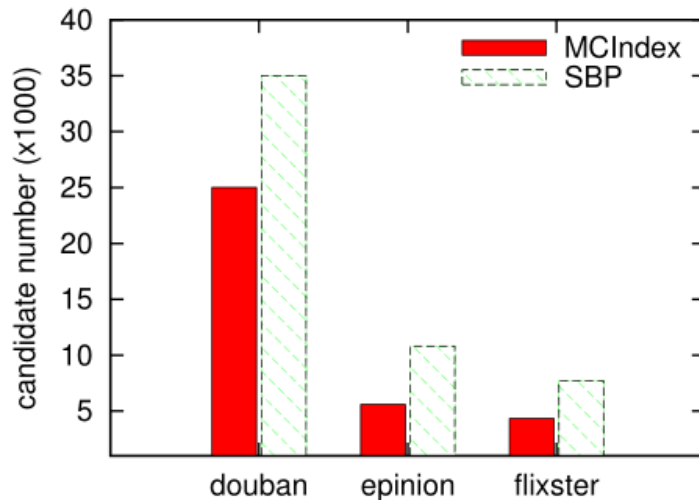
Frontiers of Computer Science, DOI: [10.1007/s11704-019-7235-0](https://doi.org/10.1007/s11704-019-7235-0)

# Problems & Ideas

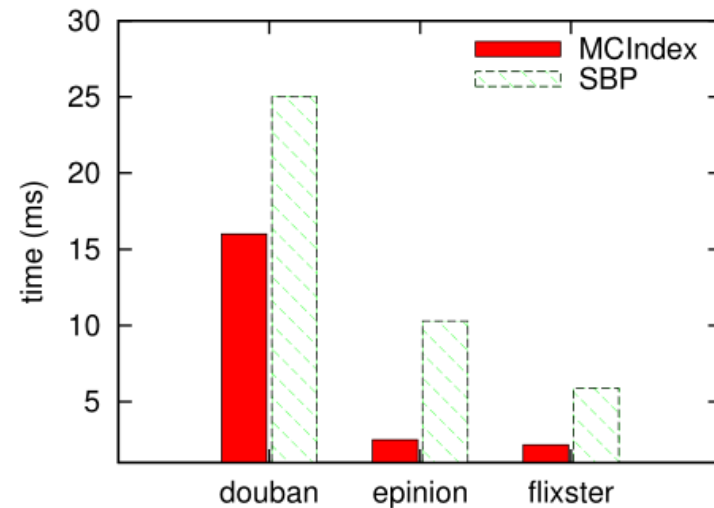
- Problem definition: Given a graph  $G$  and a set of nodes  $Q$ , find the MC  $M$  of  $G$  such that the Jaccard similarity  $\text{jac}(M, Q)$  is maximized.
- Ideas:
  - Filtering-and-verification framework
  - For filtering, to improve the efficiency
    - We construct an inverted index  $I$  of all Maximal Cliques (MCs) of  $G$
    - For each inverted list in the inverted index, we construct a novel index (MCIndex) and propose a distance-based pruning rule
  - For verification, to improve the efficiency
    - We propose a binary representation of the MCs for efficiently computing the size of the intersection with  $Q$

# Main Contributions

- The number of candidates can be ~50% of the state-of-the-art method SBP
- The query time can be ~20% of the state-of-the-art method SBP



(a) # of candidates



(b) query time