Asieh GHANBARPOUR, Khashayar NIKNAFS, Hassan NADERI, 2020. Efficient keyword search over graph-structured data based on minimal covered *r*-cliques. *Frontiers of Information Technology & Electronic Engineering*, 21(3):448-464.
https://doi.org/10.1631/FITEE.1800133

Efficient keyword search over graph-structured data based on minimal covered *r*-cliques

Keywords: Keyword search; Graph mining; Information retrieval; Database; Clique

Corresponding author: Hassan Naderi
E-mail: naderi@iust.ac.ir
ORCID: http://orcid.org/0000-0002-3296-8505

Motivation

- Keyword search is known as an alternative for structured query language to query databases.
- A result to a keyword query is a compactly connected structure which covers all or part of the queried keywords.
- Usually a large number of answers are relevant to a keyword query.
- Finding relevant answers to a given query efficiently and ranking the retrieved answers in an effective way are still two challenging problems in the keyword search domain.

Main idea

- Due to the large volume of search space and the large number of relevant results to each query, we focus on the pruning techniques, pivoting techniques, and type restrictions to reduce the exploration cost of finding answers.
- We use parallel processing techniques for efficient retrieval of answers.
- Besides trying to improve the efficiency, we aim at retrieving the top-*k* results in the order of their final ranking.

Concepts & Methods

1) We propose a new semantics for the results of a keyword search in which the keyword vertices are close to each other.

2) We present BKS as the base algorithm for retrieving all the MCC_rs of a homogenous graph, in which pruning methods are employed as in the Bron-Kerbosch algorithm. BKS uses pivoting and type restrictions to reduce the exploration cost of finding MCC_rs.

3) Algorithm BKSR is proposed as an improved version of BKS. This algorithm searches the graph more efficiently by imposing a frequency-based ordering on the graph's vertices.

Concepts & Methods (Cont'd)

4) We propose BKSM and BKSRM as distributed versions of BKS and BKSR algorithms, respectively. These algorithms rely on the distributive nature of the base algorithms for parallel search of the search space.

5) A heuristic algorithm is presented to produce the top-*k* results of a keyword query in a polynomial delay. This algorithm can incrementally generate results by observing the distance restrictions. Using this algorithm, an approximate version is proposed for each of the exact algorithms.

Major results

- Efficiency evaluation

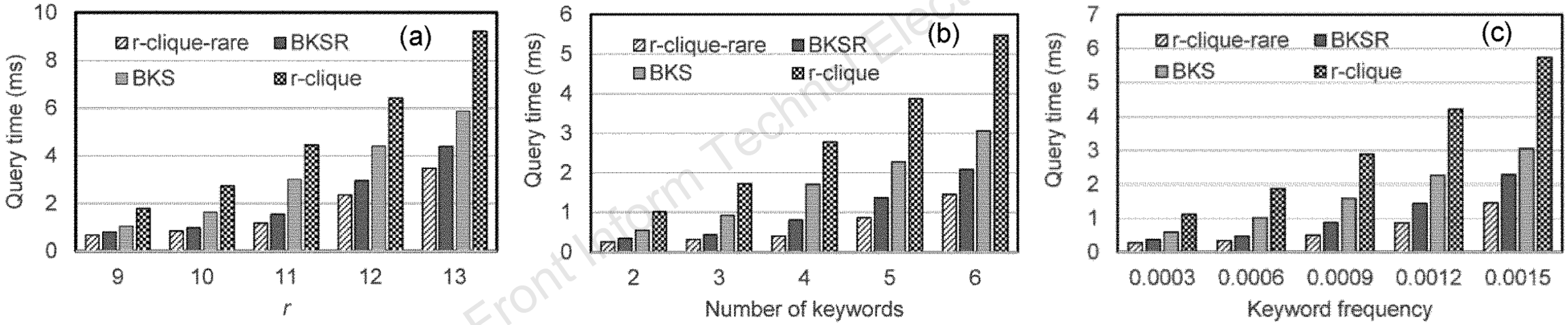
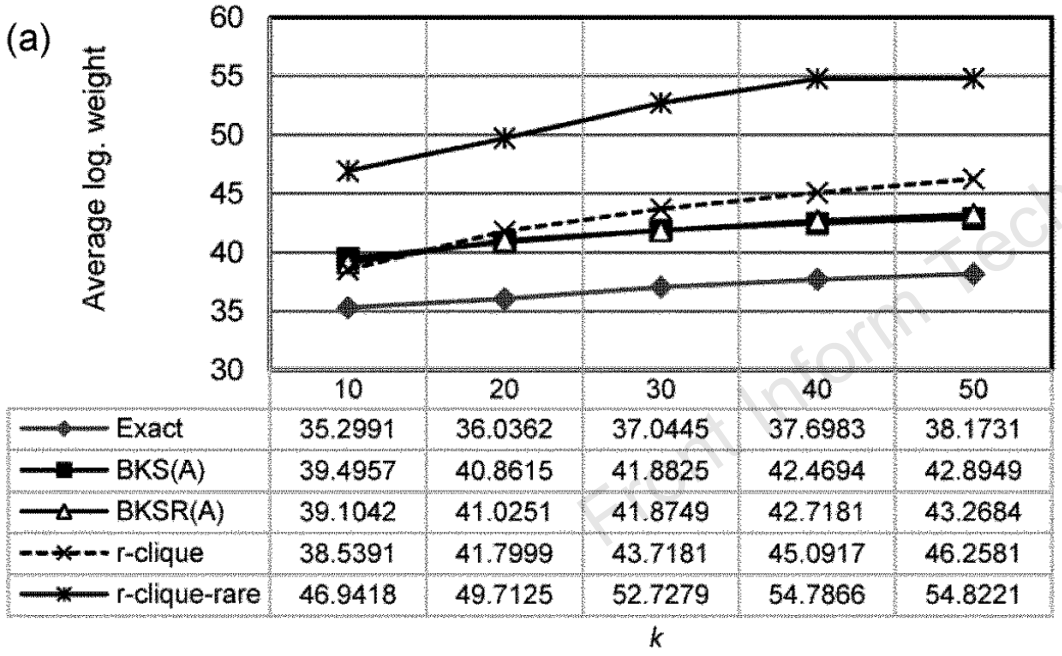


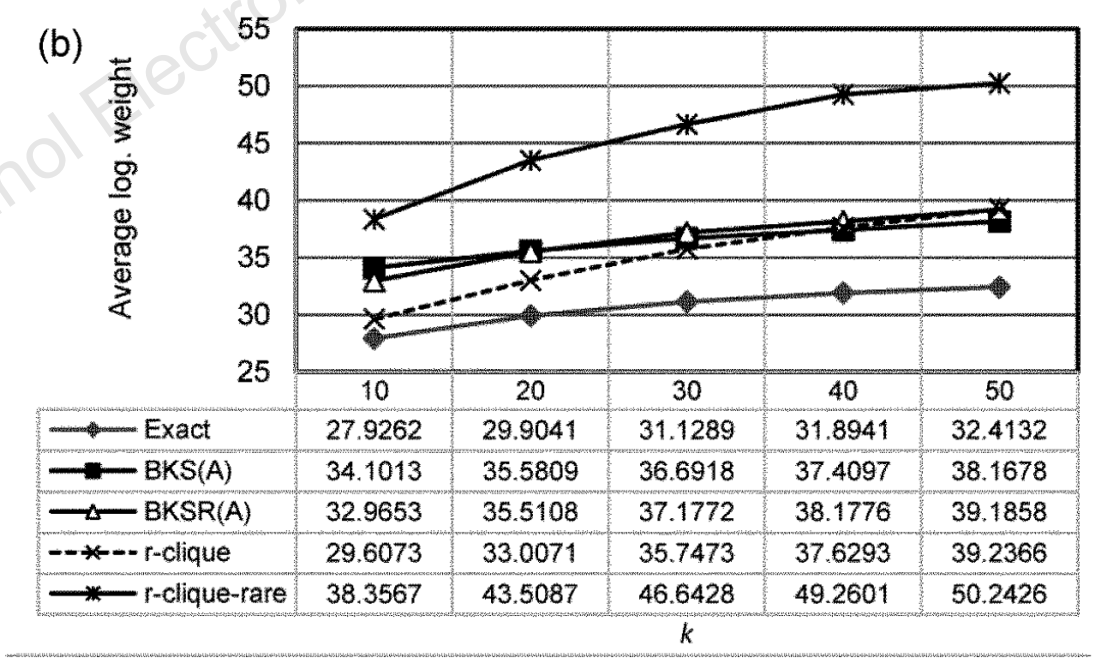
Fig. 4 Execution times of the systems based on different factors over the IMDb dataset: (a) query time vs. *r*; (b) query time vs. the number of keywords; (c) query time vs. the keyword frequency

Major results (Cont'd)

- Effectiveness evaluation



	10	20	30	40	50
Exact	35.2991	36.0362	37.0445	37.6983	38.1731
BKS(A)	39.4957	40.8615	41.8825	42.4694	42.8949
BKSR(A)	39.1042	41.0251	41.8749	42.7181	43.2684
r-clique	38.5391	41.7999	43.7181	45.0917	46.2581
r-clique-rare	46.9418	49.7125	52.7279	54.7866	54.8221



	10	20	30	40	50
Exact	27.9262	29.9041	31.1289	31.8941	32.4132
BKS(A)	34.1013	35.5809	36.6918	37.4097	38.1678
BKSR(A)	32.9653	35.5108	37.1772	38.1776	39.1858
r-clique	29.6073	33.0071	35.7473	37.6293	39.2366
r-clique-rare	38.3567	43.5087	46.6428	49.2601	50.2426

Fig. 9 The average logarithmic weight of results when querying the keywords with frequency 0.0003 (a) and 0.0015 (b)

Conclusions

- We have proposed a group of scalable branch and bound algorithms which retrieve MCC_r as the results of a keyword query. These algorithms employ different strategies of pruning to reduce the time of answering a keyword query.
- The time complexity of the proposed algorithms shows, in theory, the better performance of the proposed algorithms in comparison to algorithms in the previous works.
- In the effectiveness evaluation based on a compactness measure, the proposed algorithms significantly outperformed those in the previous works in terms of generating results with maximum distance *r*.