Keywords-Driven and Weight-aware Paper Recommendation via Paper Correlation Pattern Mining

Hanwen Liu (hanqingliu2019@outlook.com)
Nanjing University of Science and Technology https://orcid.org/0000-0002-2905-370X

Jun Hou
Nanjing Vocational University of Industry Technology

Qianmu Li
Nanjing University of Science and Technology

Jian Jiang
Nanjing University of Science and Technology

Research

Keywords: paper recommendation, correlation pattern mining, query keyword, weighted paper correlation graph

Posted Date: January 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-144551/v1

Abstract

Currently, readers often prefer to search for their interested papers based on a set of typed query keywords. As the keywords of a paper is often limited, paper recommender systems often need to recommend a set of papers which collectively satisfy the readers' keyword query. However, the topics of recommended papers are probably not correlated with each other, which fail to meet the readers' requirements on in-depth and continuous academic research. Furthermore, although existing paper citation graphs can model the papers' correlations, they often face the data sparse problem which blocks accurate paper recommendations. To address these issues, we propose a keywords-driven and weight-aware paper recommendation approach, named LP-PRk+w (link prediction-paper recommendation), based on a weighted paper correlation graph. Concretely, we firstly optimize the existing paper citation graph modes by introducing a weighted similarity, after which we obtain a weighted paper correlation graph. Then we recommend a set of correlated papers based on the weighted paper correlation graph and the query keywords from readers. At last, we conduct large-scale experiments on a real-world Hep-Th dataset. Experimental results demonstrate that our proposal can improve the paper recommendation performances considerably, compared to other related solutions.

Full Text

This preprint is available for download as a PDF.

Tables

Table 1. Symbol definition

Symbol	Definition
p / v	A paper
$k_1, \ldots, k_z$	The paper contains keywords
	A correlation relationship
Q / K	A set of query keywords
/	A set of authors / keywords
	A set of nodes
	A set of edges
	A set of weights
	W-PCG
S_k	An inverted index
$T_w(Q)$	An optimal answer tree
Q^1 / Q^2	Two queues
$T_{wmin}(v, K')$	A minimum group weighted Steiner trees rooted at v
R_p	Recommendation result

Table 2. The α, β and λ = 0.3 are employed in the link prediction to obtain the number of new edges.

β \ α	0.3	0.5	0.7	0.9
0.3	190	192	136	160
0.5	264	270	276	238
0.7	258	258	264	310
0.9	258	258	258	264

Table 3. The α, β and λ = 0.9 are employed in the link prediction to obtain the number of new edges.

β\α	0.3	0.5	0.7	0.9
0.3	208	152	160	168
0.5	348	234	216	176
0.7	342	342	314	226
0.9	288	288	288	294

Figures

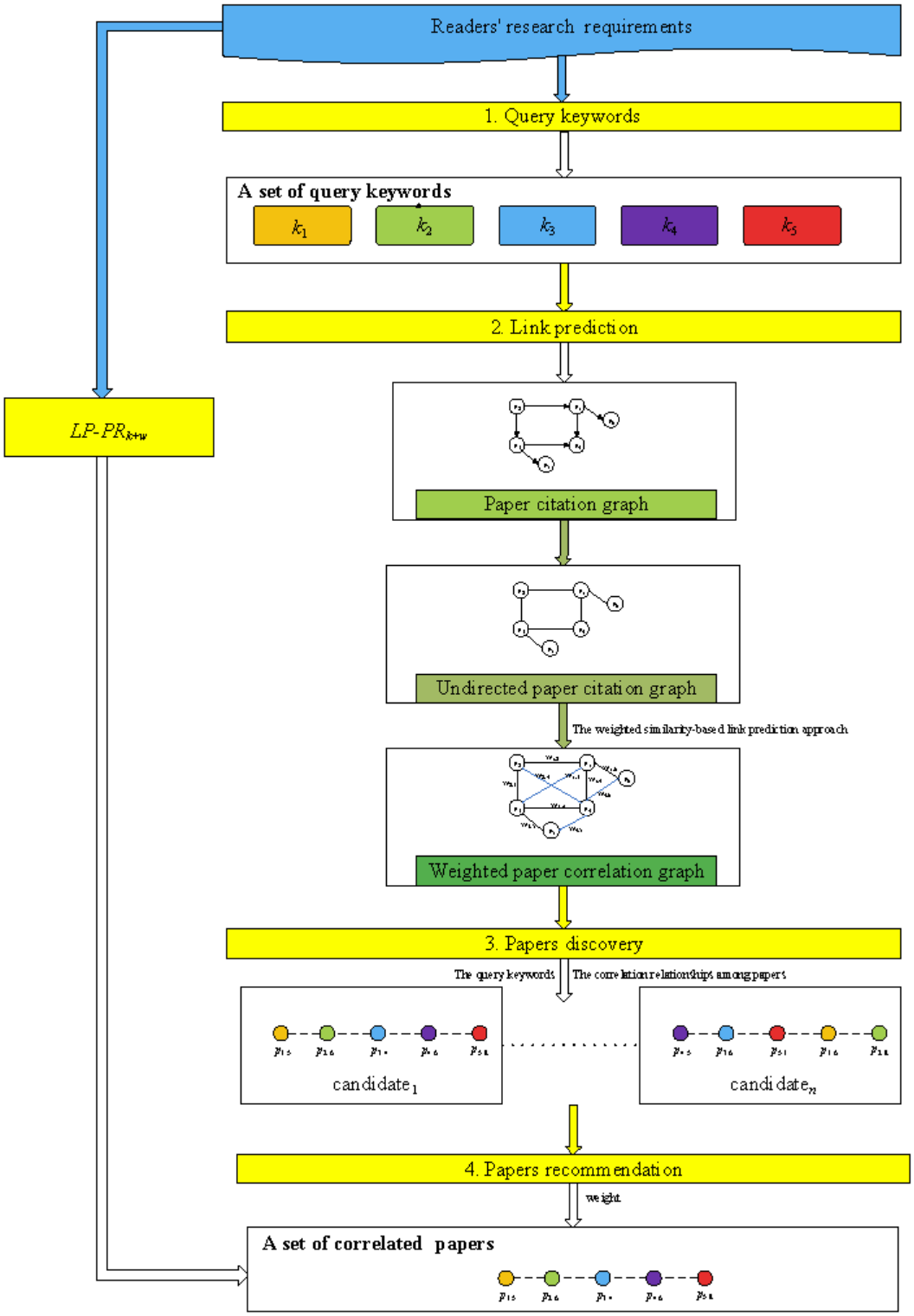


Figure 1

The recommendation process of LP-PRk+w.

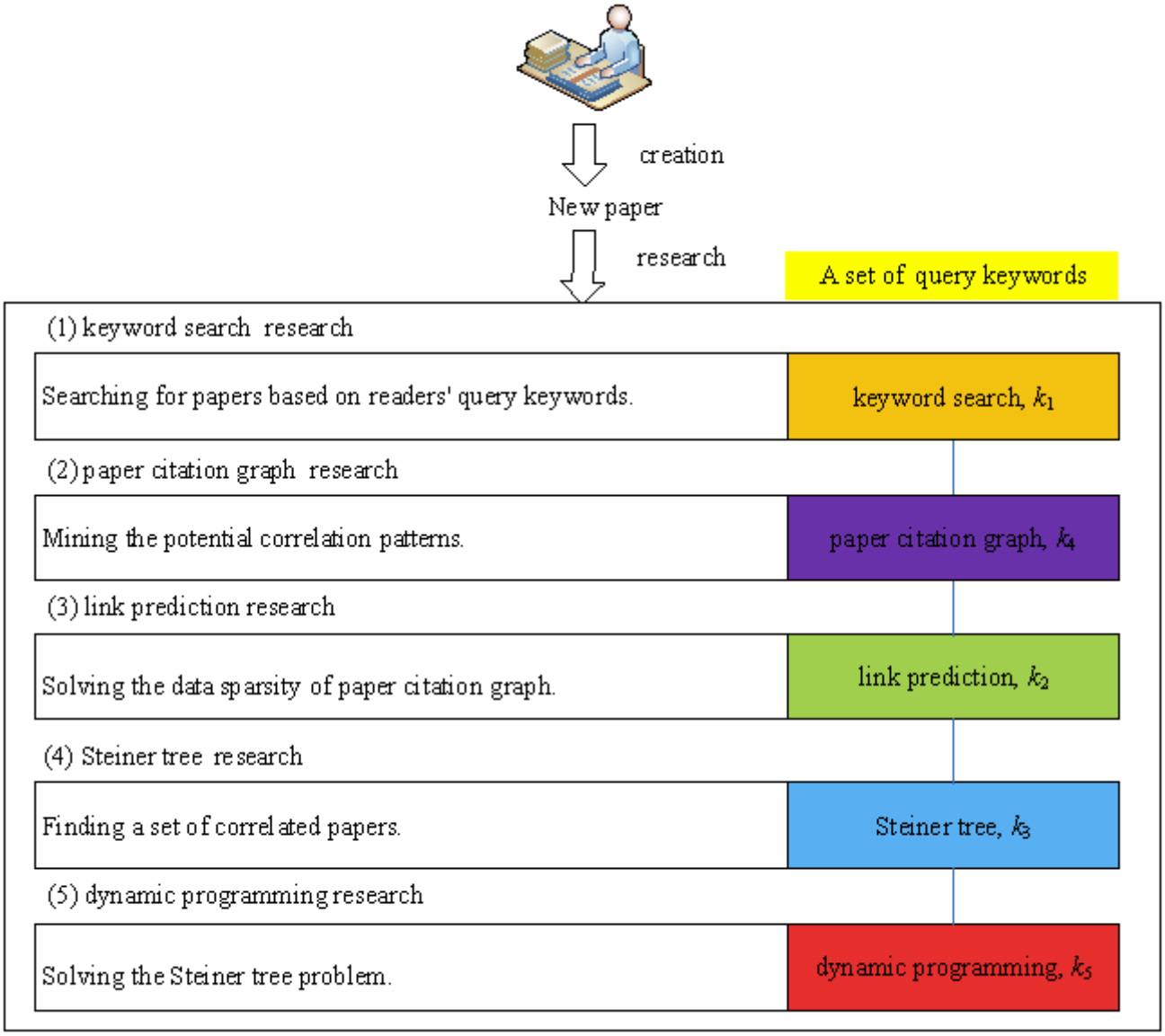


Figure 2

The paper research and creation tasks: an example.

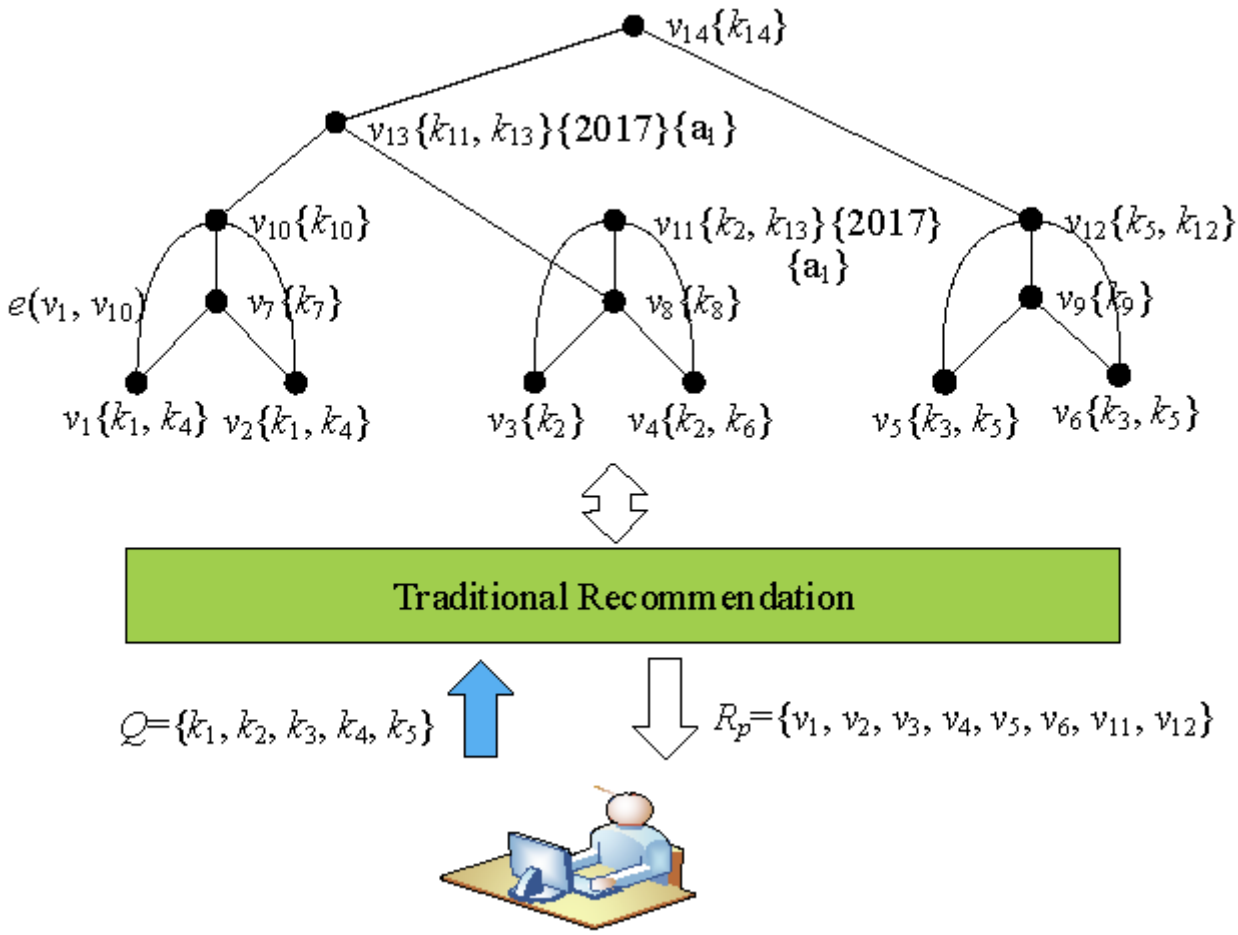


Figure 3

Tradition recommendation: an example.

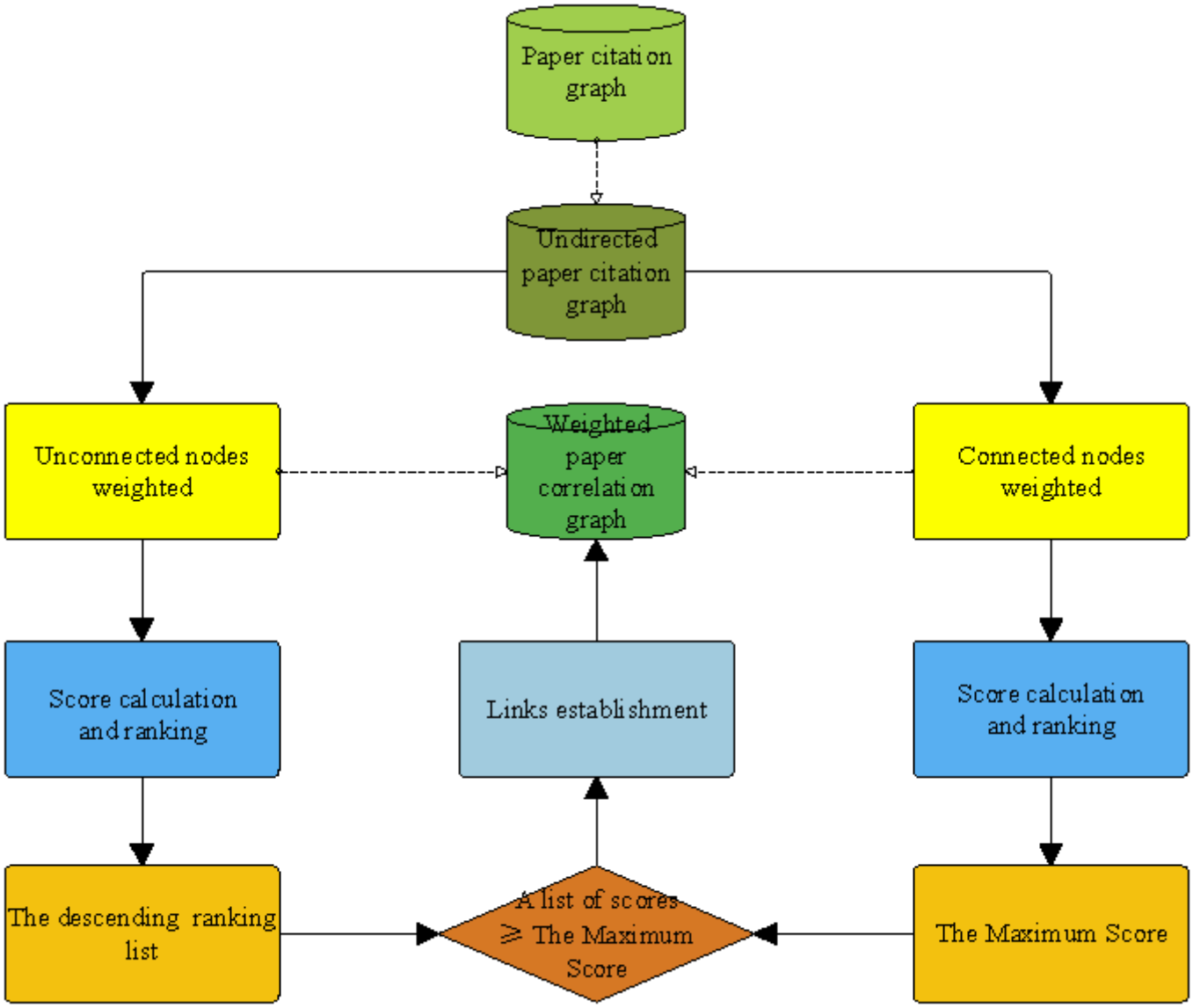


Figure 4

Process for the weighted similarity-based link prediction: overview.

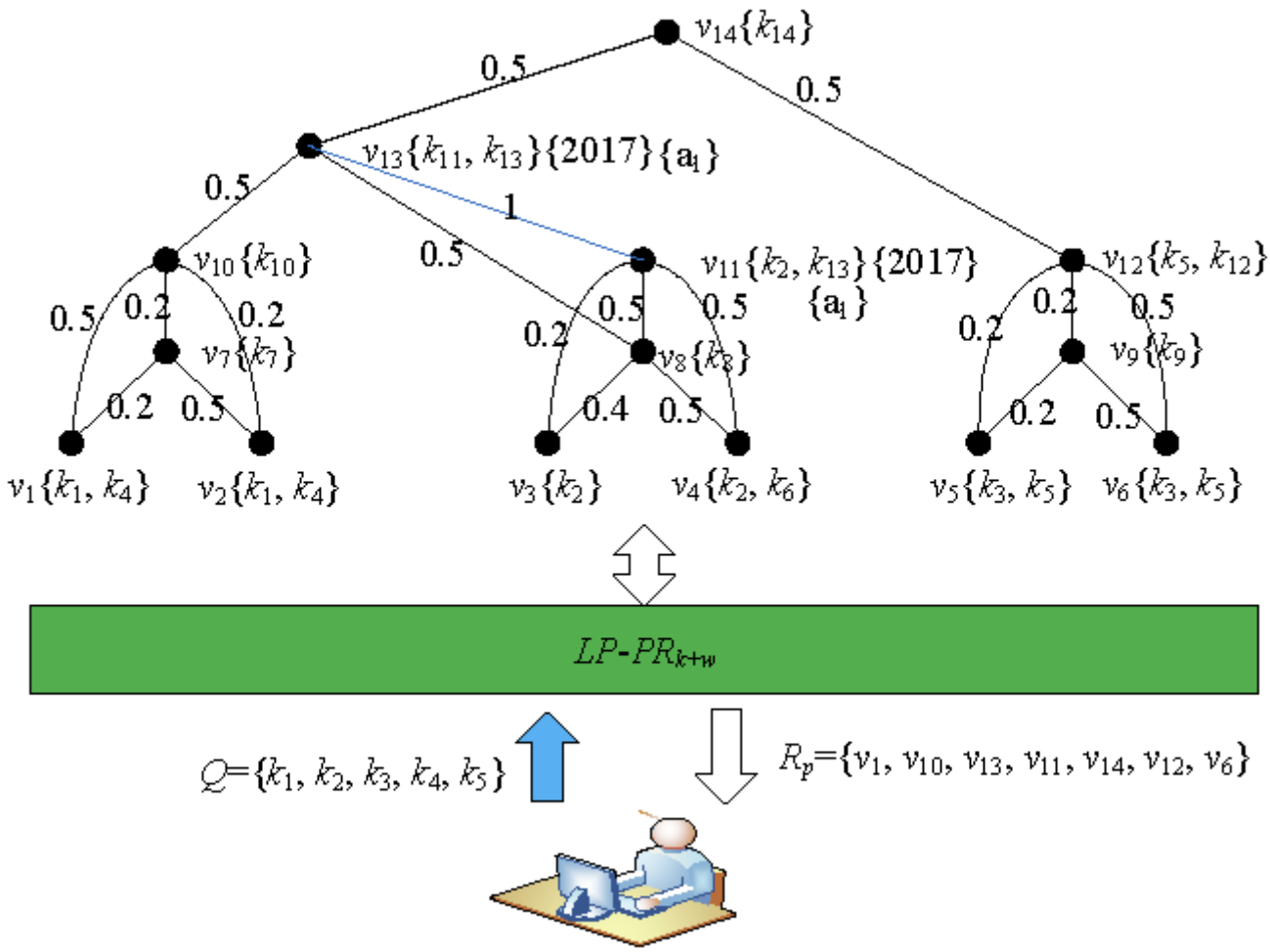


Figure 5

LP-PRk+w: the same paper recommendation example.

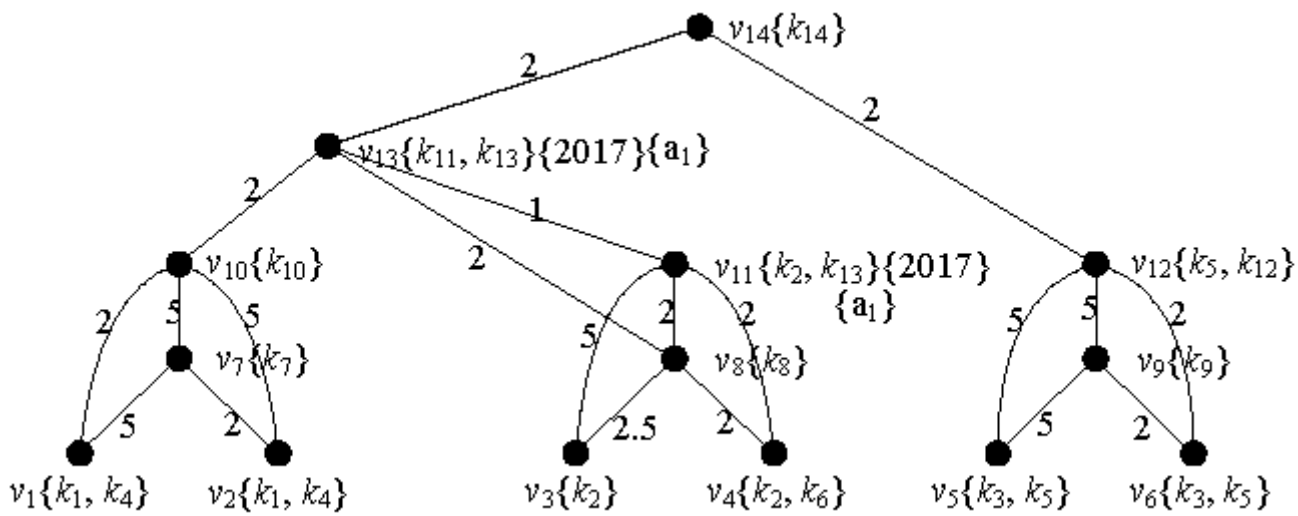


Figure 6

The W-PCG converted from Figure. 5.

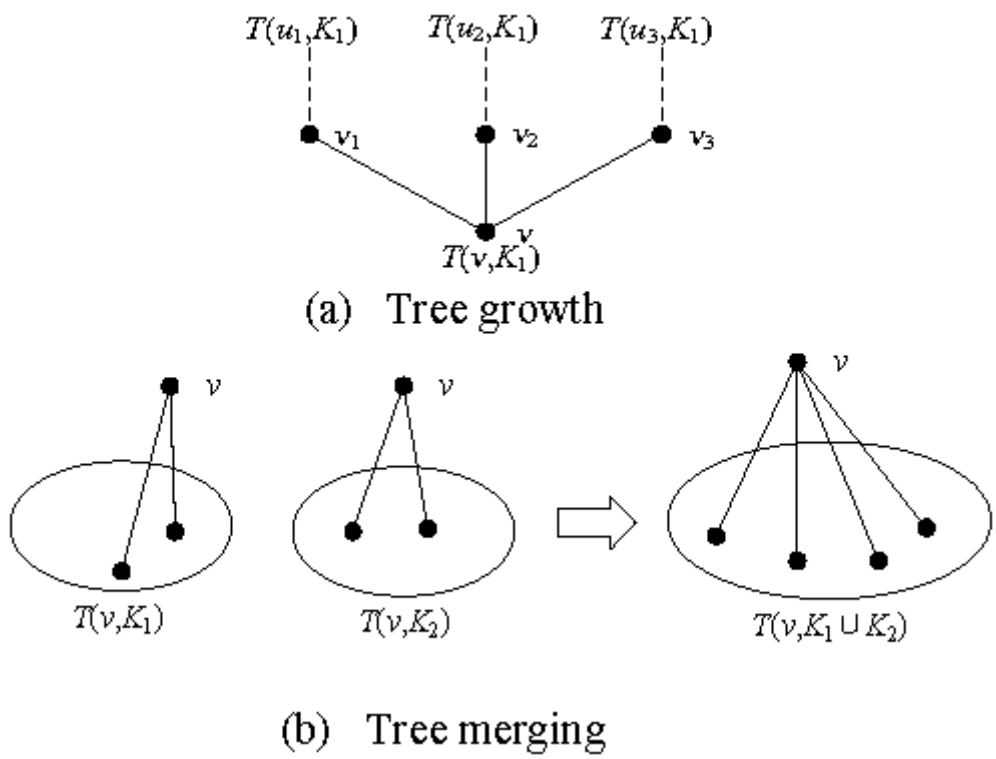
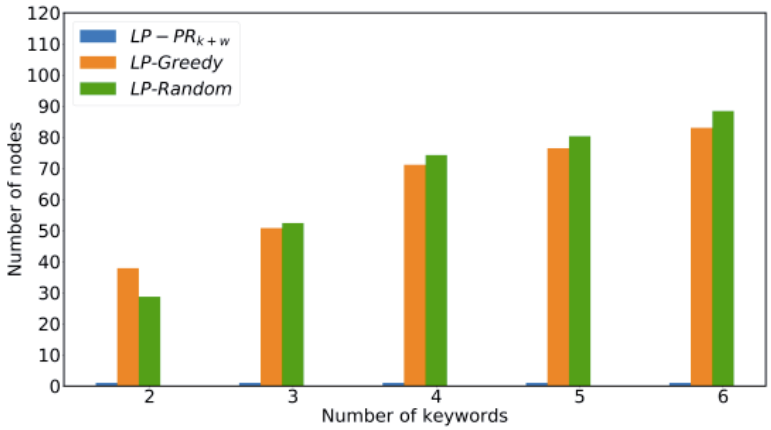
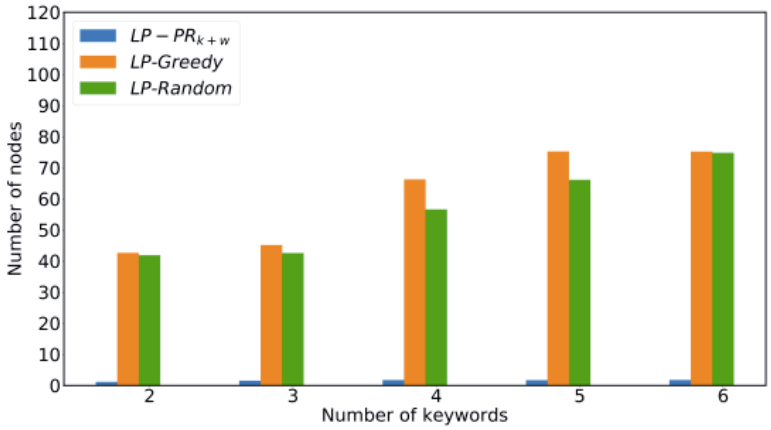


Figure 7

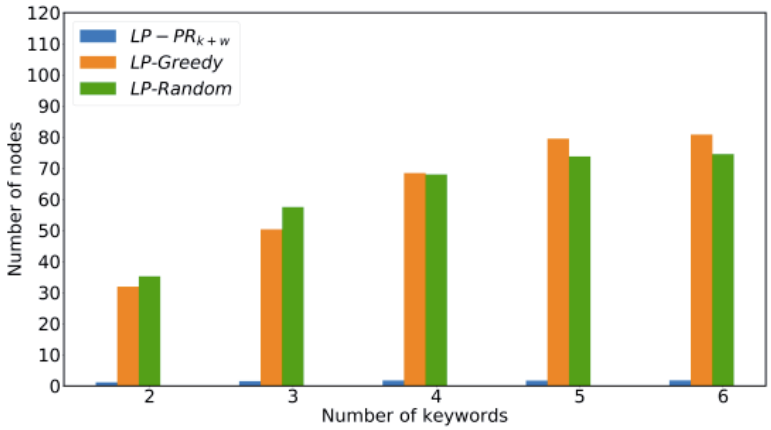
Tree operations



(a) The number of recommended nodes in set A.



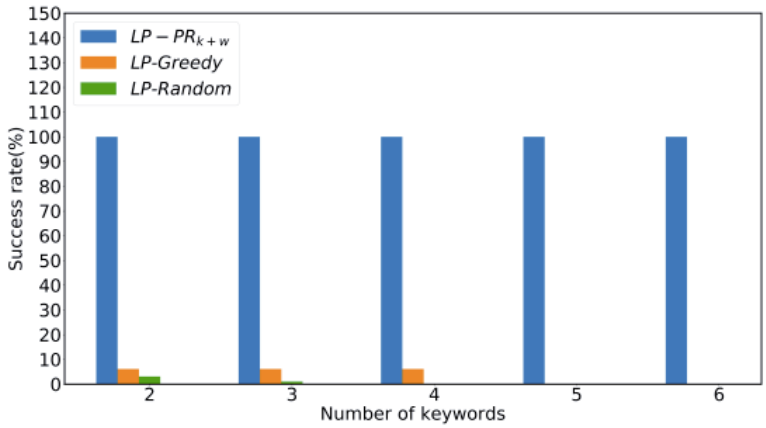
(b) The number of recommended nodes in set B.



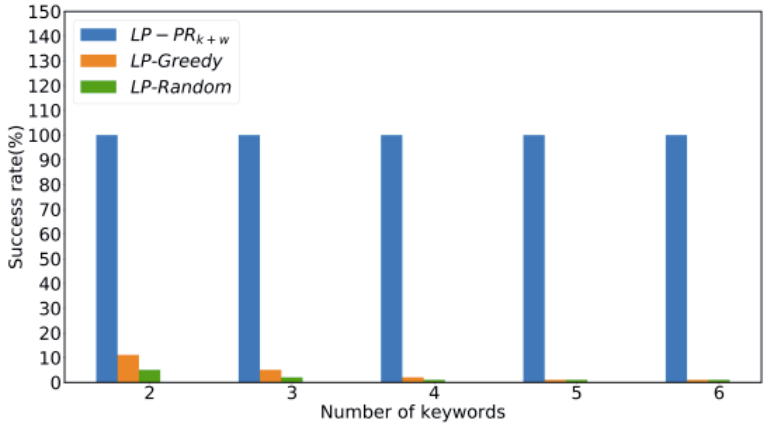
(c) The number of recommended nodes in set C.

Figure 8

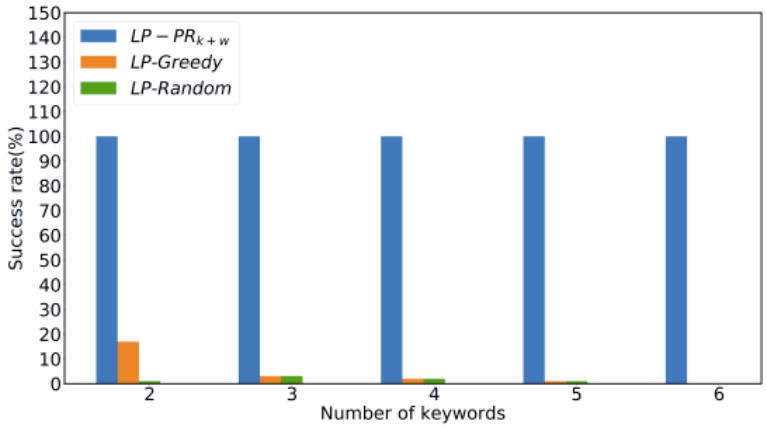
The number of recommended nodes of different approaches.



(a) The success rate in set A.



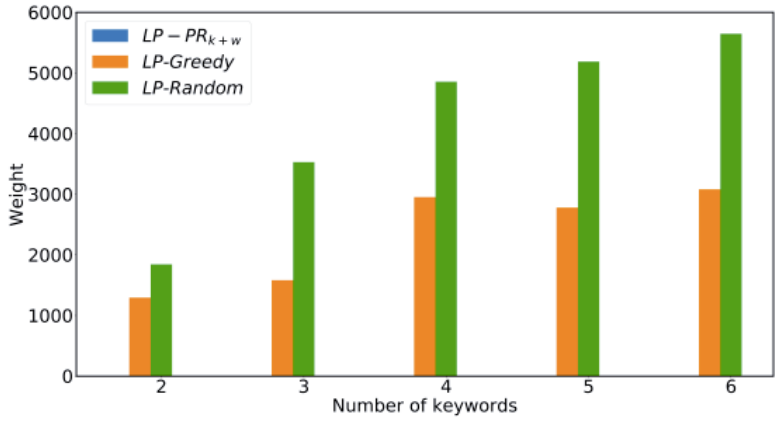
(b) The success rate in set B.



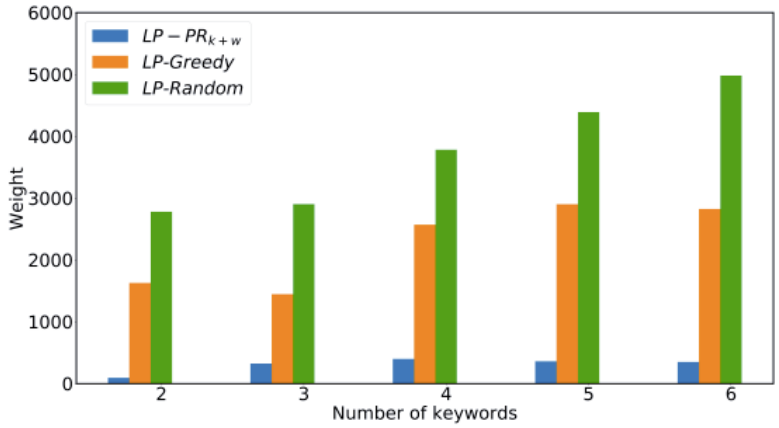
(c) The success rate in set C.

Figure 9

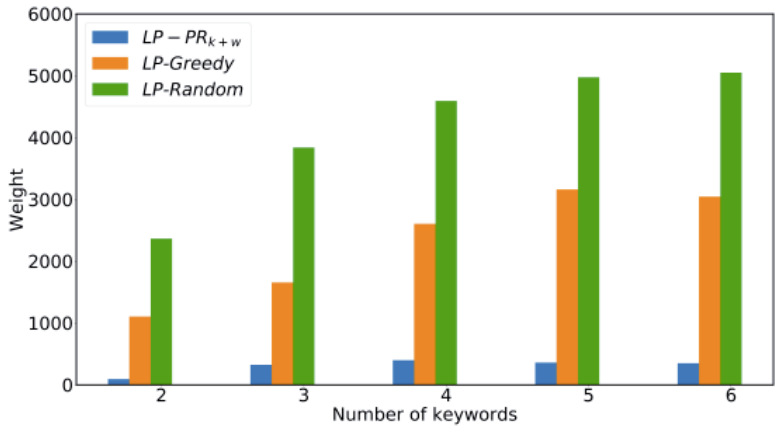
The success rate of different approaches.



(a) The weight in set A.



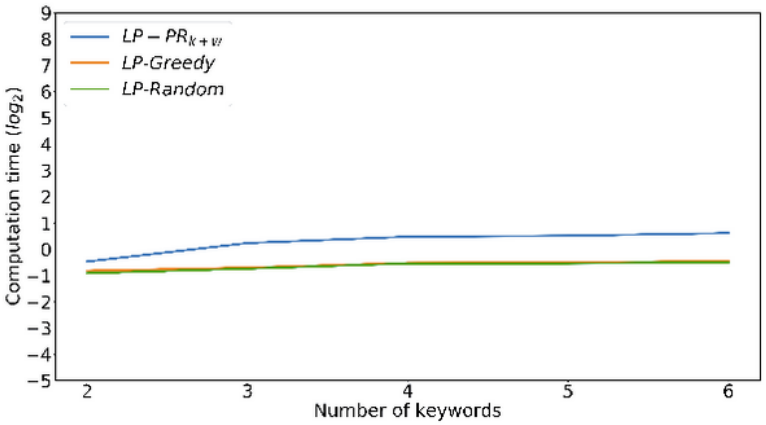
(b) The weight in set B.



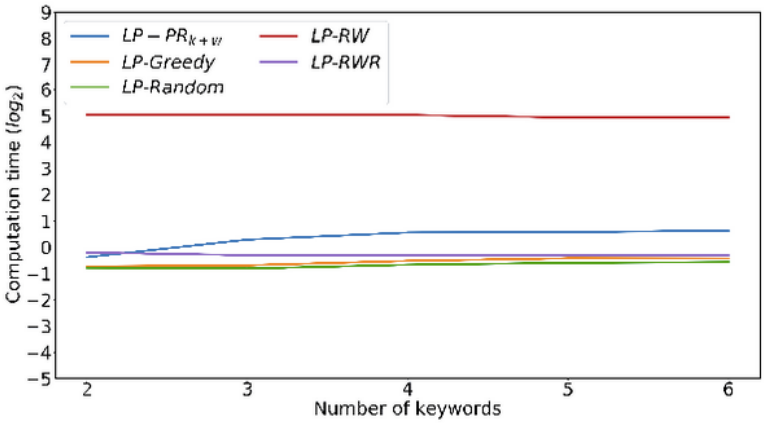
(c) The weight in set C.

Figure 10

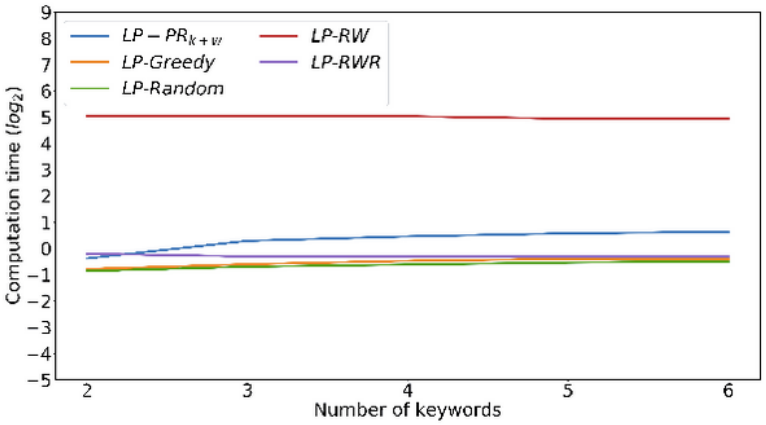
The weight of different approaches.



(a) The computation time in set A.



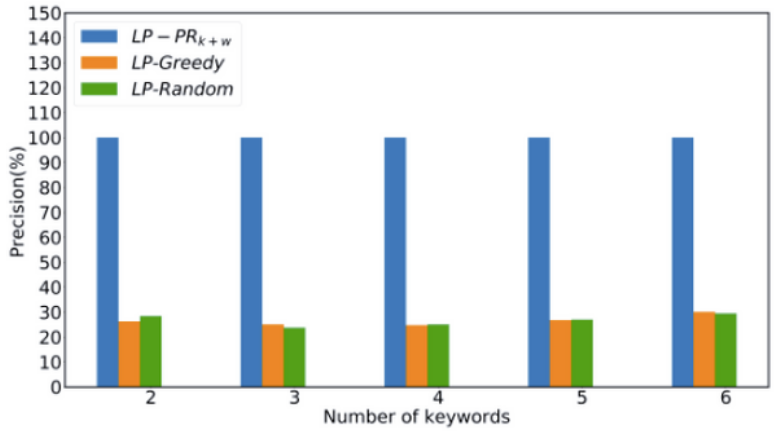
(b) The computation time in set B.



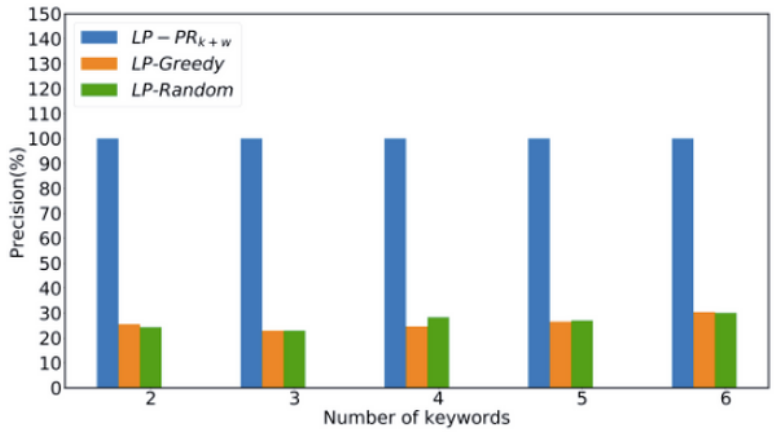
(c) The computation time in set C.

Figure 11

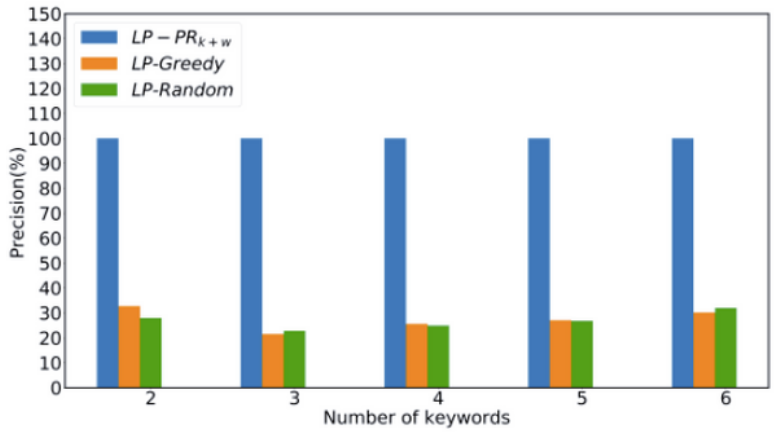
The computation time of different approaches.



(a) The precision in set A.



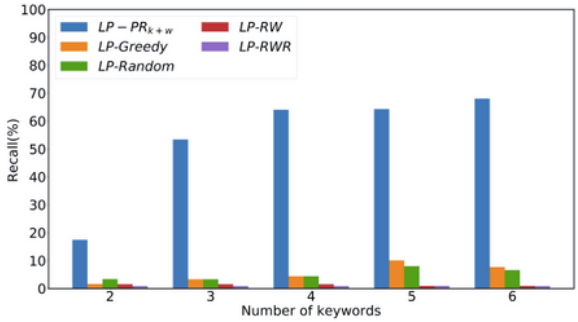
(b) The precision in set B.



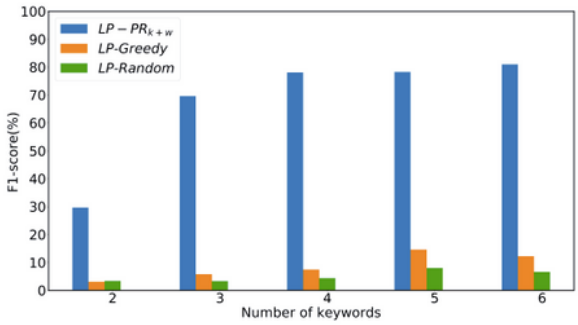
(c) The precision in set C.

Figure 12

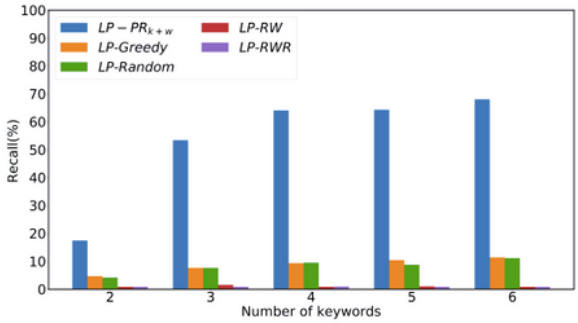
The precision of different approaches.



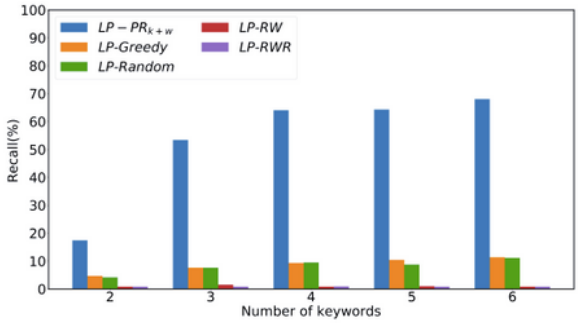
(a) The recall in set B.



(b) The F1-score in set B.



(c) The recall in set C.



(d) The F1-score in set C.

Figure 13

The recall and F1-score of different approaches.