

基于 r -子团最小覆盖的图结构数据高效关键字搜索

Asieh GHANBARPOUR^{1,2}, Khashayar NIKNAFS¹, Hassan NADERI¹

¹ 伊朗科技大学计算机工程学院, 伊朗德黑兰, 13114-16846

² 锡斯坦-俾路支斯坦大学计算机工程系, 伊朗扎黑丹, 98167-45845

摘要: 对图结构数据的查询, 关键字搜索是结构化查询语言的一种替代方式。关键字查询的结果是图结构数据上一个连接的结构, 其覆盖所有或部分关键字。文本覆盖率和结构紧凑性是评价关键字查询结果是否相关的两个主要属性。以往研究通过排序函数对比所有候选结果的上述属性。但是, 该过程耗时长, 不符合人们在交互系统中短时得到返回结果的需求。近期研究通过在搜索过程中限制搜索结果的结构形状并考察其覆盖率和紧凑性来解决上述问题。然而, 这些方法仍无法解决检索结果中存在冗余节点的问题。本文针对关键字查询结果提出基于 r -子团最小覆盖 (minimal covered r -clique, MCC_r) 的概念, 作为现有定义的扩展模型, 并给出高效算法以检测给定查询的 MCC_r 。这些算法的优势在于不仅可以检索出某个关键字查询的全部非重复 MCC_r , 还可以分布式方式执行。此外, 提出这些算法的近似版本, 以多项式时间复杂度检索最高的 k 个近似 MCC_r 。论文表明近似算法可基于成对近似排序检索出结果。基于两个真实数据集的大量实验验证了所提算法的效率和有效性。

关键词: 关键字搜索; 图挖掘; 信息检索; 数据库; 子团

<https://doi.org/10.1631/FITEE.1800133>