基本信息

姓名 孙亚辉
邮箱 yahuisun@outlook.com
微信/手机 17800199087
现在的职位 副教授, 计算机科学与技术系, 信息学院, 中国人民大学

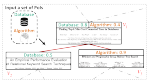
教育与工作经历

2022- **副教授, 中国人民大学, 北京 (Renmin University of China, Beijing)**
2021-2022 **讲师, 中国人民大学, 北京 (Renmin University of China, Beijing)**
2020-2021 **研究人员 (Research fellow), 新加坡国立大学 (National University of Singapore, Singapore)**
2019-2020 **研究人员 (Research fellow), 新加坡南洋理工大学 (Nanyang Technological University, Singapore)**
2018-2019 **博士后 (Postdoctoral fellow), 澳大利亚国立大学 (Australian National University, Australia)**
2014-2018 **博士, 澳大利亚墨尔本大学 (University of Melbourne, Australia)**
2012-2014 **硕士, 哈尔滨工业大学, 哈尔滨 (Harbin Institute of Technology, Harbin)**
2008-2012 **本科, 哈尔滨工业大学, 哈尔滨 (Harbin Institute of Technology, Harbin)**

研究方向

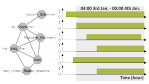
图、向量数据管理算法与系统；GPU高性能计算；图+AI；图+隐私保护，应用场景包括大模型检索增强生成、知识图谱、社交网络等（我正在寻找优秀的本科生和研究生加入课题组，鼓励同学们根据兴趣选择工业界导向的工程类、或学术界导向的算法类研究课题）。

五篇代表作



Shuang Yang (课题组学生), **Yahui Sun*** (独立通讯作者), Jiesong Liu, Xiaokui Xiao, Rong-Hua Li, and Zhewei Wei. "Approximating Probabilistic Group Steiner Trees in Graphs", **Proceedings of the VLDB Endowment** (2023) (数据库与数据挖掘领域顶会/刊； CCF A类) [**PDF**]

Consider an edge-weighted graph, and a number of properties of interests (PoIs). Each vertex has a probability of exhibiting each PoI. The joint probability that a set of vertices exhibits a PoI is the probability that this set contains at least one vertex that exhibits this PoI. The probabilistic group Steiner tree problem is to find a tree such that (i) for each PoI, the joint probability that the set of vertices in this tree exhibits this PoI is no smaller than a threshold value, e.g., 0.97; and (ii) the total weight of edges in this tree is the minimum. Solving this problem is useful for mining various graphs with uncertain vertex properties, but is NP-hard. To meet this challenge, we propose 3 approximation algorithms for solving the above problem.



Yahui Sun, Shuai Ma, and Bin Cui. "Hunting temporal bumps in graphs with dynamic vertex properties", **Proceedings of the 2022 ACM SIGMOD international conference on management of data** (2022) (数据库与数据挖掘领域顶会； CCF A类) [**PDF**]

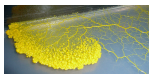
Given a time interval and a graph where vertices exhibit a property of interest (PoI) dynamically, an interesting question is: where (i.e., which part of the graph) and when (i.e., which time sub-interval) does the PoI occur frequently? To our knowledge, no work has been done to answer this question to date. We address this issue in this paper. First, we propose two approximation algorithms. Then, we propose two heuristic algorithms that produce similar solutions with, and are considerably faster than, the two approximation algorithms. Experiments on real datasets show that, in comparison with baselines built using related existing techniques, our algorithms are more suitable for answering the aforementioned question.

Yahui Sun, Xiaokui Xiao, Bin Cui, Saman Halgamuge, Theodoros Lappas, and Jun Luo. "Finding group Steiner trees in graphs with both vertex and edge weights", **Proceedings of the VLDB Endowment** (2021) (数据库与数据挖掘领域顶会/刊； CCF A类) [**PDF**]

Finding group Steiner trees is a standard approach to information retrieval in relational databases. Most existing work focuses on finding group Steiner trees in vertex-unweighted graphs, and not enough work has been done to find group Steiner trees in graphs with both vertex and edge weights. Here, we develop several algorithms to address this issue. Initially, we extend two algorithms from vertex-unweighted graphs to vertex- and edge-weighted graphs. Then, we develop several new approximation algorithms, one of which provides the tightest polynomial-time approximation guarantee to date. Experiments show that, while no algorithm is the best in all cases, our algorithms considerably outperform the state of the art in many scenarios.

Yahui Sun, Jun Luo, Theodoros Lappas, Xiaokui Xiao, and Bin Cui. "Hunting multiple bumps in graphs", **Proceedings of the VLDB Endowment** (2020) (数据库与数据挖掘领域顶会/刊； CCF A类) [**PDF**]

Bump hunting is a graph-related anomaly detection approach. A single bump is hunted in an unweighted graph in the previous work. We extend the previous work by hunting multiple bumps in a weighted graph. We prove that our extended problem can be transformed to a recently formulated prize-collecting Steiner forest problem. We further prove that this problem is NP-hard even in trees. Subsequently, we propose a fast approximation algorithm for solving this problem in trees. Based on this algorithm, we improve the state-of-the-art approximation algorithm for solving this problem in graphs. Experiments on real datasets show the dominance of our improvement over the state-of-the-art algorithms for hunting multiple bumps in graphs.



Yahui Sun, Daniel Rehfeldt, Marcus Brazil, Doreen Thomas, and Saman Halgamuge. "A Physarum-inspired algorithm for minimum-cost relay node placement in wireless sensor networks", **IEEE/ACM Transactions on Networking** (2020) (计算机网络领域顶刊； CCF A类) [**PDF**]

Relay node placement is essential in minimizing the costs of wireless sensor networks. Here, we focus on minimum-cost relay node placement. By considering the heterogeneous production and placement costs of relay nodes, our work extends the previous work that considers the costs of relay nodes to be homogeneous. Initially, we conduct some theoretical analyses on the emerging Physarum-inspired algorithms to reveal their potential of computing efficient networks. Based on these analyses, we propose an algorithm for minimum-cost relay node placement. In comparison with the state of the art, our algorithm designs wireless sensor networks with lower relay costs and similar qualities of service. Our work is particularly useful in budget-limited scenarios.

其他发表文章

Yahui Sun, Marcus Brazil, Doreen Thomas, and Saman Halgamuge. "The fast heuristic algorithms and post-processing techniques to design large and low-cost communication networks", IEEE/ACM Transactions on Networking (2019) (计算机网络领域顶刊； CCF A类) [**PDF**]

Yahui Sun, and Saman Halgamuge. "Minimum-cost heterogeneous node placement in wireless sensor networks." IEEE Access (2019). (投这篇文章的时候我和我博士导师Saman Halgamuge并不太了解IEEE Access; 这是我的, 也是Saman的, 唯一一篇IEEE Access文章；这是一篇经得起时间检验的文章) [**PDF**]

Yahui Sun, Chenkai Ma, and Saman Halgamuge. "The node-weighted Steiner tree approach to identify elements of cancer-related signaling pathways." International Conference on Bioinformatics (2017). [**PDF**]

Yahui Sun, Pathima Nusrath Hameed, Karin Verspoor, and Saman Halgamuge. "A physarum-inspired prize-collecting Steiner tree approach to identify subnetworks for drug repositioning." International Conference on Bioinformatics (2016). [**PDF**]

Yahui Sun, and Saman Halgamuge. "Fast algorithms inspired by physarum polycephalum for node weighted Steiner tree problem with multiple terminals." In IEEE Congress on Evolutionary Computation (2016) [**PDF**]

Yahui Sun, Yunhai Geng, and Shuang Wang. "Analysis and calibration of star sensor's image plane displacement." Infrared and Laser Engineering (2014). [**PDF**]

Yahui Sun, Yingying Xiao, and Yunhai Geng. "On-orbit calibration of star sensor based on a new lens distortion model." In Proceedings of the 32nd Chinese Control Conference (2013). [**PDF**]

教学经历

程序设计2荣誉课程（大一下） 2022 -
数据结构与算法2（大三上） 2022 -
程序设计实践（大二下） 2023 -
科研早培课程（大二下） 2023-

科研项目

国家自然科学基金青年基金项目 2023-2025 （负责人）
中国人民大学新教师启动基金项目 2022-2023 （负责人）
国家重点研发计划 2023-2026 （主要参与人）（负责人：李国良，题目："基于新型硬件的原生数据库系统"; 中国人民大学陈红老师负责"面向混合负载的智能异构计算技术"子课题，任务是研发基于CPU/GPU的关系、图、向量数据库，本人负责带队研发其中的基于CPU/GPU的图数据库）
国家自然科学基金重点项目 2023-2026 （主要参与人）