



基于属性图的社区搜索模式及其分类体系

赵丹枫, 孔万仔, 黄冬梅, 刘国华

引用本文:

赵丹枫, 孔万仔, 黄冬梅, 刘国华. 基于属性图的社区搜索模式及其分类体系[J]. *智能系统学报*, 2024, 19(4): 791–806.

ZHAO Danfeng, KONG Wanzai, HUANG Dongmei, et al. Community search schemata and their classification systems based on attribute graphs[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(4): 791–806.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202306050>

您可能感兴趣的其他文章

弱标记不完备决策系统的增量式属性约简算法

An incremental attribute reduction algorithm for incomplete decision system with weak labeling
智能系统学报. 2020, 15(6): 1079–1090 <https://dx.doi.org/10.11992/tis.202001017>

利用二元拟阵Kn图的一种建格方法

A constructive method of lattice using the Kn diagram of binary matroid
智能系统学报. 2017, 12(3): 333–340 <https://dx.doi.org/10.11992/tis.201704022>

基于稠密子图的社区发现算法

Community detection algorithm based on dense subgraphs
智能系统学报. 2016, 11(3): 426–432 <https://dx.doi.org/10.11992/tis.201603045>

基于权值最大圈的概念格构造算法

An algorithm for concept lattice construction based on maximum cycles of weight values
智能系统学报. 2016, 11(4): 519–525 <https://dx.doi.org/10.11992/tis.201606006>

数据偏序结构关系中的知识发现可视化方法

A novel knowledge discovery visualization method based on data partial ordered structure
智能系统学报. 2016, 11(4): 475–480 <https://dx.doi.org/10.11992/tis.201606019>

面向用户兴趣与社区关系的微博话题检测方法

Micro-blog topic detection based on users' interests and communities
智能系统学报. 2016, 11(3): 294–300 <https://dx.doi.org/10.11992/tis.201603341>

DOI: 10.11992/tis.202306050

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20240407.1236.002>

基于属性图的社区搜索模式及其分类体系

赵丹枫¹, 孔万仔¹, 黄冬梅², 刘国华³

(1. 上海海洋大学 信息学院, 上海 201306; 2. 上海电力大学, 上海 200090; 3. 东华大学 计算机科学与技术学院, 上海 201620)

摘要: 当前在属性图中的社区搜索方法较多、类型繁杂, 没有系统的分类方式, 约束了社区搜索的应用。为明确属性图社区搜索的类别, 对属性图社区搜索分类方法进行研究。首先, 首次提出属性图社区搜索模式的概念, 深入分析属性图社区搜索模式之间存在的联系, 提出属性图社区搜索模式的等价、从属、交叉、全异 4 种关系; 其次, 以搜索模式的输入图属性、输出图拓扑结构和各属性图社区搜索模式的实际意义为基础, 构建两层分类体系, 第 1 层是由输入属性图相同的模式集合构成的集族, 这里的输入属性图包括时序、空间、关键字、权值、空属性图, 第 2 层是由输出图拓扑结构及实际意义定位到的每一个具体的属性图社区搜索模式; 然后, 针对第 2 层中每一种模式, 给出对应社区搜索算法的对比分析结果; 最后, 对所有属性图社区搜索模式的特性集中分析。总体而言, 属性图社区搜索模式不仅为理解和分析复杂网络结构提供有力工具, 也为解决实际问题提供新的视角和方法。

关键词: 图论; 属性图; 社区搜索; 模式; 内聚性; 拓扑结构; 关系; 社区搜索算法**中图分类号:** TP311 **文献标志码:** A **文章编号:** 1673-4785(2024)04-0791-16

中文引用格式: 赵丹枫, 孔万仔, 黄冬梅, 等. 基于属性图的社区搜索模式及其分类体系 [J]. 智能系统学报, 2024, 19(4): 791-806.

英文引用格式: ZHAO Danfeng, KONG Wanzai, HUANG Dongmei, et al. Community search schemata and their classification systems based on attribute graphs[J]. CAAI transactions on intelligent systems, 2024, 19(4): 791-806.

Community search schemata and their classification systems based on attribute graphs

ZHAO Danfeng¹, KONG Wanzai¹, HUANG Dongmei², LIU Guohua³

(1. School of Information, Shanghai Ocean University, Shanghai 201306, China; 2. Shanghai University of Electric Power, Shanghai 200090, China; 3. School of Computer Science and Technology, Donghua University, Shanghai 201620, China)

Abstract: At present, there are many community search methods in the attribute graph, and there is no systematic classification method, which restricts the application of community search. In order to clarify the category of community search in attribute graph, the classification method of attribute community search is studied. Firstly, the concept of attribute community search schema is proposed to analyze the relationship between attribute community search schemata in depth, proposing four relationships of community search mode of attribute graph: equivalence, affiliation, intersected and exclusion. Secondly, a two-layer classification system is constructed based on the input graph attributes of the search mode, the topology of the output graph and the practical significance of the search mode of each attribute community. The first layer is a family of sets composed of the same set of schemata in the input attribute graph. The input attribute graph here includes sequence, space, keyword, weight, and empty attribute graph. The second layer is each specific community search schema located by the topology and practical meaning of the output graph. Then, the comparative analysis result of corresponding community search algorithm is given for each schema in the second layer. Finally, the characteristics of all the community search modes of attribute graphs are analyzed centrally. Overall, the attribute graph community search pattern not only provides a powerful tool for understanding and analyzing complex network structures, but also provides a new perspective and method for solving practical problems.

Keywords: graph theory; attributed graph; community search; schema; cohesiveness; topology; relation; community search algorithm

收稿日期: 2023-06-28. 网络出版日期: 2024-04-09.

基金项目: 国家自然科学基金青年基金项目(42106190); 国家自然科学基金面上项目(61972241).

通信作者: 刘国华. E-mail: ghliu@dhu.edu.cn.

图作为有效描述大数据的数据结构, 扮演着越来越重要的角色。现实世界中的许多数据及其关系都可以用属性图描述, 进而通过属性图上的

研究来解决基于大数据的实际问题。例如蛋白质网络分析^[1]、社交网络分析^[2]、知识图谱分析^[3]等。社区搜索问题由 Sozio 等^[4]于2010年提出,是近年来的图计算领域的热门研究课题。社区搜索问题的输入为图,输出为满足一定内聚性的连通子图(即社区)。研究者们从不同角度、层次对社区搜索问题进行了研究。早期社区搜索问题主要关注结构内聚性,近年来随着在应用领域的深入,属性内聚性也成为关注点之一。社区的结构内聚性是以不同的拓扑结构来约束的,经典的子图拓扑结构按照其出现顺序,可分为 k -ECC、 k -clique、 k -core 和 k -truss。属性图中的社区搜索研究关注的属性包括空属性以及节点的关键字、位置、时序、权值和边的权值属性等。

当前对社区搜索的研究,集中在提出并解决具有不同输入图属性与输出图结构组合的新的社区搜索问题,或完善现有社区搜索问题的解决方法,从而集中了大量类型繁杂的社区搜索问题。而这些社区搜索问题之间隐含的联系及可能存在的关系并没有引起重视。本文总结了社区搜索的3大要素:输入初始图的属性、输出子图的拓扑结构及各社区搜索问题具体实际意义,将各类具有共性的社区搜索问题定义成属性图社区搜索模式,并分析归纳了这些模式之间的4种关系:等价、从属、交叉、全异。这种定义及分类能够快速定位到所需要的最合适的社区搜索模式。除此之外,这些属性图社区搜索模式种类繁多且没有系统的分类方式,极大程度地限制了社区搜索的应用。本文提出了属性图社区搜索模式的两层分类体系。第1层基于各模式输入图的属性,将所有属性图社区搜索模式分为6个集族;第2层按照返回社区拓扑结构及具体实际意义,从这些集族拆出具体的属性图社区搜索模式。

本文第1节分别定义了社区搜索模式及属性图社区搜索模式,并根据4种拓扑结构定义了各属性图社区搜索模式之间的关系;第2节提出了对属性图社区搜索模式的两层分类体系,并对分别对空属性与时序、空间、关键字、权值社区搜索模式进行了具体定义及归纳,分析了现阶段属于各模式的算法,并对个模式之间的关系分析后系统表述;第3节对非空属性图社区搜索模式对比分析,直观展示各模式的重点特征,可根据需求快速选择合适的社区搜索模式,给出该使用案例并在真实数据集中运行;第4节对社区搜索模式的发展提出自己的建议,为后来研究者的深入研究提出几种的思路。

1 属性图社区搜索模式及其关系

本节形式化定义了各种属性图并根据社区搜索的3大要素(输入初始图的属性、输出子图的拓扑结构、各社区搜索问题的实际意义)将各类具有共性的社区搜索问题定义成属性图社区搜索模式,并定义了这些模式之间的4种关系。

定义1 属性图:一个属性图是 $G=(V, E, \Sigma, L)$, 其中 V 为点集, E 为边集, $(\Sigma=C) \vee (\Sigma=K) \vee (\Sigma=W)$, C 为二维坐标集, K 为关键字集, W 为权值集, L 为顶点或边对应的函数。

社区搜索模式中所用到的输入属性图可分为空属性图、时序属性图、位置属性图、关键字属性图和权值属性图。

定义2 空属性图

当一个属性图 $G_O=(V, E, \Sigma, L)$ 中 Σ 为空集且 $\forall e \in E$ 是一个二元组 (u, v) 时,称 G 为简单图或空属性图。

定义3 时序属性图

时序属性图 $G_T=(V, E, \Sigma, L)$ 是一个包含点集 V 、边集 E 的图, $\forall e \in E$ 是一个三元组 (u, v, t) 。其中 $u, v \in V$, t 是 u 和 v 交互的时间点。

定义4 位置属性图

位置属性图 $G_C=(V, E, \Sigma, L)$ 是一个包含点集 V 、边集 E 的图, $\Sigma=C$, C 中元素为二维坐标系 (v_x, v_y) , $\forall v \in V$ 与 Σ 中元素通过函数 L 一一对应。

定义5 关键字属性图

关键字属性图 $G_K=(V, E, \Sigma, L)$ 是一个包含点集 V 、边集 E 的图, $\Sigma=K$, K 为关键字集, $\forall v \in V$ 与 Σ 中元素通过函数 L 一一对应。

定义6 权值属性图

权值属性图 $G_W=(V, E, \Sigma, L)$ 是一个包含点集 V 、边集 E 的图, $\Sigma=W$, W 为权值集。 $\forall v \in V$ 与 Σ 中元素 W , 通过函数 L 一一对应时,称该权值属性图为点权属性图;当 $\forall e \in E$ 与 Σ 中元素 W_e 通过函数 L 一一对应时,称该权值属性图为边权属性图。

定义7 社区搜索^[4]

给定一个无向图 $G(V, E)$ 、一个查询节点集 $Q \subseteq V$ 、一个优度函数 f , 需要找到一个 G 的诱导子图 $H(V_H, E_H)$, 满足如下性质: 1) V_H 包含 Q ; 2) H 是连通的; 3) 所有的可行解中 $f(H)$ 值满足给定约束条件。

社区搜索模式 (community search schema, CSS) 是一类社区搜索的集合。

定义8 属性图社区搜索

给定一个属性图 $G=(V, E, \Sigma, L)$, 这里的属性图可为定义 2~6 中的空属性图、时序属性图、位置属性图、关键字属性图或权值属性图, 一个查询节点集 $Q \subseteq V$, 一个属性评分函数 f , 需要找到一个 G 的诱导子图 $H(V_H, E_H, \Sigma_H, L_H)$, 满足如下性质: 1) V_H 包含 Q ; 2) H 是连通的, 且满足一定拓扑结构; 3) 所有的可行解中 $f(H)$ 值最大或最小。

属性图社区搜索模式 (attribute community search schema, ACSS) 是一类属性图社区搜索问题的集合, 表示为 $S = \{q | \forall v \in S\}$ 为同一类属性图社区搜索问题}。

由于属性图可以分为空属性图、时序属性图、位置属性图、关键字属性图、权值属性图, 所以属性图社区搜索模式可分为空属性图社区搜索模式、时序属性图社区搜索模式、位置属性图社区搜索模式、关键字属性图社区搜索模式和权值属性图社区搜索模式。

社区搜索的输出社区需要满足一定的结构内聚性, 社区的结构内聚性是以不同的拓扑结构来约束的, 下面是 4 种常见拓扑结构。

定义 9 k -core

k -core 是一个给定图 G 的最大子图 H , 满足 $\forall v \in V, \text{deg}_H(v) \geq k$ 。

定义 10 k -truss

k -truss 是一个给定图 G 的最大子图, 且满足 $\forall e \in E_H, \text{sup}(e, H) \geq k-2$ 。

定义 11 k -clique

给定图 G 的最大子图 $H(V_H, E_H)$, 且满足任意的 $v_i, v_j \in V_H$ 之间都有一条边。

定义 12 边连通度

对于一个给定图 $G=(V, E)$, 在去掉任意 $k-1$ 条边后, 所得的子图仍然连通, 但去掉 k 条边后不连通, 则称 k 为图 G 的边连通度, 记为 $\lambda(G)$ 。

定义 13 k -ECC

k -ECC 是一个满足边连通度为 k 的 G 的子图 H 。由定义 9、10、11、13 可知^[5]:

- 1) k -clique 一定是 k -truss 的子图;
- 2) k -truss 的连通子图一定是特定 k -ECC 的子图;
- 3) k -truss 一定是 $(k-1)$ -core 的子图;
- 4) k -ECC 一定是 k -core 的子图。

所以, 上述 4 种结构的内聚度大致为: $k\text{-core} \leq k\text{-ECC} \leq k\text{-truss} \leq k\text{-clique}$ 。

属性图社区搜索模式最重要的 3 点要素为输入初始图的属性、输出社区的拓扑结构和各社区搜索问题的具体实际意义。下文根据前 2 种要素为社区搜索模式(下称“模式”)提出 4 种社区搜索

模式关系(下称“关系”), 包括等价、从属、交叉、全异。

定义 14 等价关系

若模式 A 与模式 B 中所有问题的输入初始图的属性、输出社区的拓扑结构都相同, 则称模式 A 与模式 B 具有等价关系, 那么在没有特殊情况下, 模式 A、B 中的算法都可以解决模式 A 和 B 中的问题, 如 2.3.1 节中模式 19 与模式 21。

定义 15 从属关系

若模式 A 与 B 中所有问题的输入初始图的属性相同, 且模式 A 与模式 B 中所有问题对应返回社区的拓扑结构分别为 k -clique 和 k -truss、 k -clique 和 k -core、 k -truss 和 k -core 或 k -ECC 和 k -core 时, 则称模式 B 从属于模式 A。无特殊情况下, 模式 B 所返回的社区为模式 A 的可行解, 如 2.3.2 节中模式 25 与模式 31。

定义 16 交叉关系

若模式 A 与模式 B 中所有问题的输入初始图的属性相同, 输出社区的拓扑结构之间对应关系不属于定义 15 中提到的 4 种关系, 则称模式 A 与模式 B 具有交叉关系。无特殊情况下, 模式 A 所返回的社区与模式 B 所返回的社区存在一定的关联, 如 2.3.2 节中模式 22 与模式 24。

定义 17 全异关系

若模式 A 与模式 B 中所有问题的输入初始图的属性不同, 则称模式 A 与模式 B 全异。

2 两层分类体系

属性图社区搜索模式的 3 要素为初始输入属性图的属性、输出子图的拓扑结构以及各属性图社区搜索问题的实际意义, 本节根据这 3 要素提出了属性图社区搜索模式的两层分类体系。依据定义 2~6, 本文将输入属性图的属性作为第 1 层分类依据将各属性图社区搜索模式分成输入图属性为时序、位置、关键字、点权、边权和空属性的集族。以输出子图的拓扑结构及属性图社区搜索模式的实际意义作为第 2 层分类依据, 将第 1 层分出的 6 个集族再次具体分为各类具体的属性图社区搜索模式, 具体见定义 18、25、29、31、34。

如图 1 所示, 所有属性图社区搜索模式在第 1 层中被划分为 6 个集族, 其中权值属性图社区搜索模式按照输入属性图的权值为顶点的权值或是边的权值可进一步细分为点权以及边权属性图中的社区搜索模式。第 2 层中除空属性图社区搜索模式外的 4 类属性图搜索模式按照返回社区的

拓扑结构及模式的实际意义将其分为各个具体的社区搜索模式;由于空属性图社区搜索模式大多

不存在实际意义,按照返回社区的拓扑结构分为各个具体的社区搜索模式。

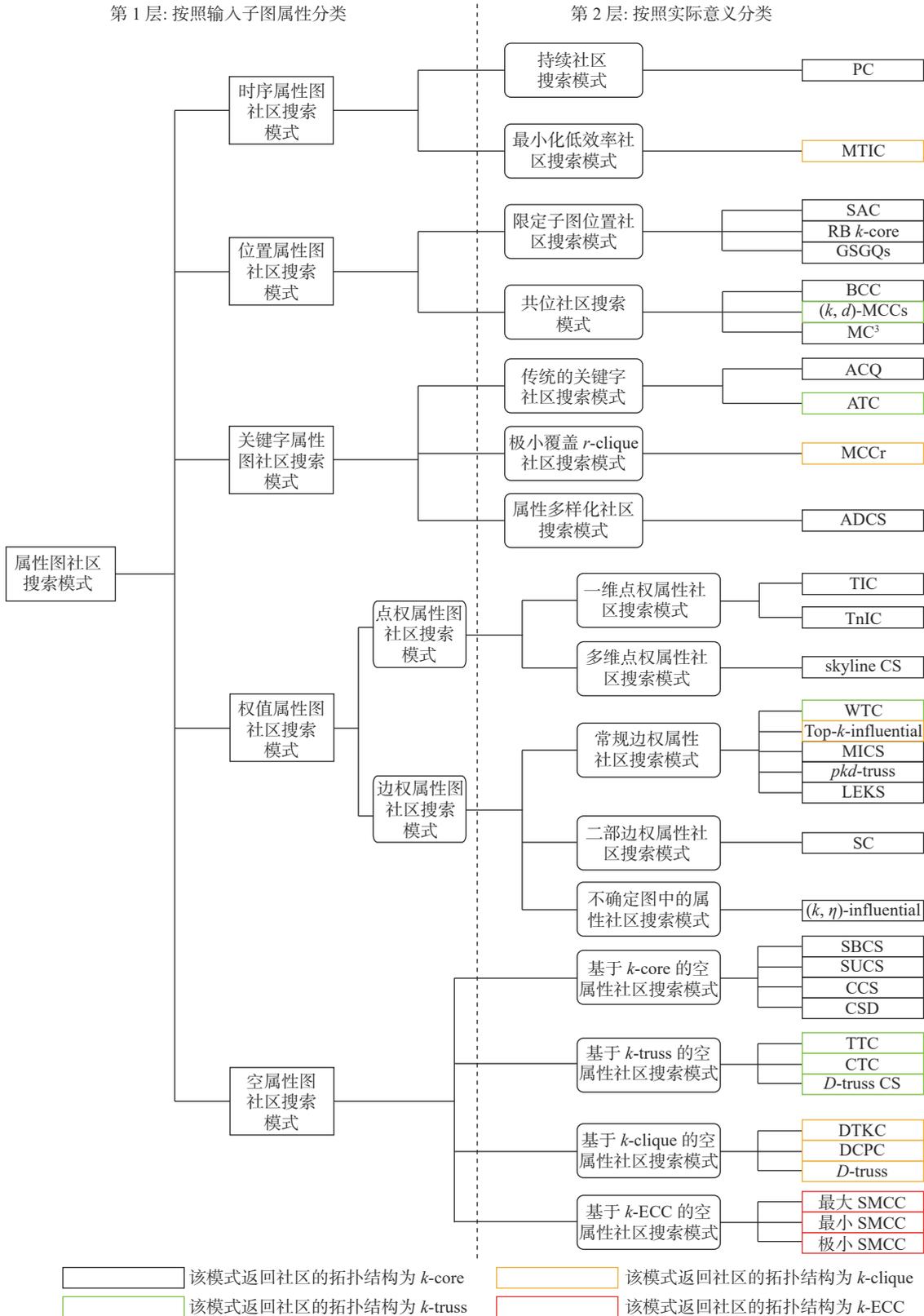


图 1 两层分类体系

Fig. 1 Two-tier classification system

2.1 空属性图社区搜索模式

早期的社区搜索问题研究很少考虑在实际中

的应用,初始输入图多为简单图。本节基于这些简单图上的社区搜索问题提出了空属性图社区搜

索模式,并按照返回社区拓扑结构的不同将其划分。

定义 18 空属性图社区搜索模式

给定一个空属性图 $G_O=(V, E, \Sigma, L)$ 、一个查询节点集 $Q \subseteq V$ 、一个属性评分函数 f , 需要找到一个 G 的诱导子图 $H(V_H, E_H, \Sigma_H, L_H)$, 满足如下性质: 1) V_H 包含 Q ; 2) H 是连通的, 且满足一定拓扑结构; 3) 所有的可行解中 $f(H)$ 值最大。

2.1.1 基于 k -core 的空属性图社区搜索模式

模式 1 规模无限制的社区搜索^[4] (size unbounded community search, SBCS) 模式

给定一个空属性图 $G_O=(V, E, \Sigma, L)$ 、一个查询节点集 $Q \subseteq V$ 、一个优度函数 f , 返回一个 G_O 的子图 $H(V_H, E_H)$, 满足以下性质: 1) $Q \subseteq V_H$; 2) H 是连通的; 3) 在所有可行解 H_i 中 H 具有最大的最小度。

SBCS 模式包含全局搜索算法、局部搜索算法以及基于 ShellStruct 索引结构和 CoreStruct 索引结果的搜索算法^[6] 共 3 种算法。SBCS 模式应用在大型图上返回的结果往往不太理想

模式 2 规模有限制的社区搜索^[4] (size bounded community search, SUCS) 模式

给定 1 个空属性图 $G_O=(V, E, \Sigma, L)$, 1 个查询节点集 $Q \subseteq V$, 2 个整数 d, k , 1 个优度函数 f , 返回 1 个 G_O 的子图 $H(V_H, E_H)$, 满足性质: 1) $Q \subseteq V_H$; 2) H 是连通的; 3) $D_Q(H) \leq d$ ^[4]; 4) $|V_H| \leq k$; 5) 所有的可行解中, H 的得分 $f(H)$ 最大。

该社区搜索模式包括 GREEDYDIST 和 GREEDYFAST 共 2 种启发式算法^[4]。类似的, Cui 等^[7] 提出了具有阈值约束的社区搜索问题。

模式 3 条件社区搜索 (condition community search, CCS) 模式^[8]

给定连通图 $G=(V, E)$ 和节点集 $V' \subset V$, 寻找至少一个满足如下条件的节点集 H : 1) $H \subset V$; 2) $G[H]$ 是连通的, $G[H]$ 表示 H 的诱导子图; 3) H 满足定义在 V' 上的用户给定搜索条件 F ; 4) H 满足用户给定的社区定义 C 。

对于一个条件社区搜索模式, 如果有且仅有一组节点变量的取值能满足它的搜索条件 F , 则称其为单项条件社区搜索 (single condition community search, SCCS)。为了解决 SCCS 模式, Zhu 等^[8] 给出了“社区搜索+过滤”方法, 即用查询节点进行社区搜索、用禁止节点进行过滤以及基于标签传播给点加权的方法。

定义 19 (k, l) -core

给定 1 个有向图 $G=(V, E)$, 2 个非负数 k, l 。 (k, l) -core 是一个满足入度小于给定 k 值、出度大

于给定 k 值的 G 的最大子图。

模式 4 有向图中社区搜索 (community search on directed graphs, CSD) 模式^[9]

给定 1 个有向空属性图 $G_O=(V, E, \Sigma, L)$ 和 2 个非负整数 k, l 以及 1 个查询节点 q , 返回 1 个连通子图 $G_Q \subseteq G_O$, G_Q 是 1 个 (k, l) -core。

文献 [9] 中利用 (k, l) -cores 的嵌套特性, 给出了 3 种方法可以优化时间及空间复杂度。

2.1.2 基于 k -truss 的空属性图社区搜索模式

定义 20 三角连通

在图 $G=(V, E)$ 中给定 2 个三角形 Δ_1, Δ_2 。如果 2 个三角形有一条公共边, 则称这 2 个三角形为邻接的。对 $\forall e_1, e_2 \in E$, 如果 e_1, e_2 属于同一个三角形或通过若干个邻接的三角形可以相互到达, 则称 e_1, e_2 是三角连通的。

定义 21 Degree-Support

给定图 G 一个子图 $H(V_H, E_H)$ 、一个非负数 α , 则边 $e(u, v) \in E_H$ 的 Degree-Support 定义为 $\text{deg}_{\text{sup}_H}(e) = \max\{\text{sup}_H(e) + 2\alpha \cdot \text{sup}_H(e)\}$ 。

定义 22 k -core-truss

给定图 G 、一个非负参数 α 、一个整数 $k \geq 2$ 、一个 k -core-truss H 是 G 的极大子图, 且对于 H 中每条边 e , 都满足 $\text{deg}_{\text{sup}_H}(e) \geq k$ 。

Liu 等^[10] 研究了在有向图中的社区搜索模式并提出了 D -truss 拓扑结构, 并在该结构的基础上提出了 D -truss 社区搜索模式。

定义 23 D -truss

给定一个有向图 G 、2 个整数 k_c, k_f 满足 $\text{c}_{\text{sup}_H}(e) \geq k_c$ 且 $\text{f}_{\text{sup}_H}(e) \geq k_f$ 的 G 的极大子图 H 为 D -truss, 也叫做 (k_c, k_f) -truss。

本节根据以上拓扑结构将在空属性图中基于 k -truss 的社区搜索模式分为三角连通 truss 社区搜索模式、基于 k -truss-core 的社区搜索模式^[11] 和基于 D -truss 的社区搜索模式^[10]。

模式 5 三角连通 truss 社区搜索 (triangle-connected truss community, TTC) 模式^[11]

给定一个空属性图 $G_O=(V, E, \Sigma, L)$ 、一个查询节点 $q \subseteq V$ 、一个整数 $k \geq 2$, 返回所有满足下面条件的子图 $H \subseteq G_O$: 1) 结构内聚性, H 包含查询节点 q , 且 $\forall e \in E_H, \text{sup}(e, H) \geq (k-2)$; 2) 三角连通性, $\forall e_1, e_2 \in E_H$, 都是三角连通的; 3) 极大子图, H 是满足上述 2 条性质的极大子图。

TTC 搜索模式包含在线搜索算法、基于 k -truss 索引的算法和一种基于 TCP-index 的搜索算法^[12] 以及基于 EquiTruss-index^[13] 的搜索算法。

模式 6 最密集社区 (closest truss community,

CTC)搜索模式^[14]

给定一个图 G 和一个查询节点集 Q , 返回一个子图 $H \subseteq G$, 满足: 1) H 包含 Q 和一个拥有最大 k 值的连通的 k -truss; 2) H 是满足性质 1 且具有最小直径(图中任意 2 点间距离的最大值)的 G 的子图。

CTC 模式中性质 1 要求社区中包含查询节点集 Q 中的所有元素, 性质 2 规定图直径来约束社区中所有节点的接近程度, 有效地避免了搭便车 (free rider) 效应。

模式 7 D -truss 社区搜索模式

给定一个有向图 $G=(V_G, E_G)$ 和 2 个整数 k_c, k_f , 一个查询节点集 $Q \subseteq V_G$ 。 D -truss 社区搜索旨在找到一个弱连通子图 $H \subseteq G$ 满足: 1) $Q \subseteq V(H)$; 2) H 是一个 D -truss; 3) $diam(H)$ 是满足上述 2 条性质的子图 H_i 中最小的。

D -truss 社区搜索模式对应全局搜索算法、局部搜索算法以及基于 D -truss 索引的算法共 3 种算法^[11]。

2.1.3 基于 k -clique 的空属性图社区搜索模式

模式 8 多样化 Top- k clique 搜索 (diversified top- k Clique, DTKC) 模式^[15]

给定空属性图 $G_O=(V, E, \Sigma, L)$ 、一个整数 k 。 DTKC 搜索模式旨在找到一个集合 H 满足: 1) $\forall C \in H$ 是一个极大团; 2) $|H| < k$; 3) $|cov(D)|$ 是最大的。

极大团枚举问题为 DTKC 搜索模式提供 2 个基础的解决方案: EnumAll 和 EnumSub。 该模式还对应 EnumKBasic^[15]、TOPKLS^[16] 2 个算法。

定义 24 k -clique 分量

以 C 表示一个 k -clique 中的连通分量, 那么一个 k -clique 分量就是 C 中所有 k -clique 内节点的集合。

模式 9 最密集 clique 过滤社区搜索 (densest clique percolation community, DCPC) 模式

给定一个空属性图 $G_O=(V, E, \Sigma, L)$ 和一系列的查询节点 $Q \subseteq V$, 需要找到包含 Q 中所有节点的具有最大 k 值的 k -clique 分量。

DCPC 的一种解决方案是用 k -clique 分量检测算法^[17] 来检查是否存在 k -clique 分量。 为了有效地支持在线 DCPC 搜索, Yuan 等^[18] 先后提出了团邻接树、有序邻接树 2 种树结构, 并由此提出了基于 DCPC-index 的搜索算法, 能够更加快速地查询 KCPC。

2.1.4 基于 k -ECC 的空属性图社区搜索模式

模式 10 Steiner 最大连通子图 (steiner maximum connected component, SMCC) 搜索模式^[19]

给定一个空属性图 $G_O=(V, E, \Sigma, L)$ 、一组查询节点 $Q \subseteq V$, 返回 G_O 的子图 $H(V_H, E_H)$ 满足以下性

质: 1) V_H 包含 q ; 2) 在所有满足 k -ECC 结构的 H_i 中, H 的边连通度 $\lambda(H)$ 是最大的; 3) 不存在其他子图 H' 满足上述性质且有 $H \subseteq H'$ 。

模式 11 最小 SMCC 搜索模式^[20]

给定一个空属性图 $G_O=(V, E, \Sigma, L)$ 、一组查询节点 $Q \subseteq V$, 返回 G_O 的子图 $H(V_H, E_H)$, H 是一个 SMCC 且在所有满足条件的 H_i 中 $|V_H|$ 最小。

模式 12 极小 SMCC 搜索模式^[20]

给定一个空属性图 $G_O=(V, E, \Sigma, L)$ 、一组查询节点 $Q \subseteq V, H'(V', E')$, H' 是一个 SMCC, 要求在图 G_O 中没有其他 H' 是一个 SMCC, 且满足 $H' \subset H$ 。

如表 1, 本文对不同子图的内聚性、使用次数、典型在线搜索算法时间复杂度进行了对比。 通过对近年来社区搜索方向的 30 多种模式分析可以发现, 学者们对于最终返回社区的子图结构的选择有着一致的偏向性。 学者们更倾向于选择 k -core 结构进行社区搜索研究。 在表 1 中: n 为节点数量, m 为边数量, s 为极大团个数, T 为枚举极大团的时间, h, l 为常数界。

表 1 不同拓扑结构对比
Table 1 Comparison of different topology

结构	内聚性	使用度	在线搜索算法时间复杂度
k -core	▲	▲▲▲▲	$O(n)$
k -truss	▲▲▲	▲▲	$O(m^{1.5})$
k -clique	▲▲▲▲	▲▲	$O(\log n T)$
k -ECC	▲▲	▲	$O(hlm)$

注: ▲ 数目越多代表该度量越高。

2.2 时序、位置属性图社区搜索模式

时间、空间属性是图数据中常见的属性, 具有这样属性的属性图被称为时序属性图和位置属性图。

2.2.1 时序属性图中的社区搜索模式

时序属性图即边具有时间间隔的属性图。

定义 25 时序属性图社区搜索模式

给定一个时序属性图 $G_T=(V, E, \Sigma, L)$ 、一个查询节点集 $Q \subseteq V$ 、一个属性评分函数 f , 需要找到一个 G 的诱导子图 $H(V_H, E_H, \Sigma_H, L_H)$, 满足性质: 1) V_H 包含 Q ; 2) H 是连通的, 且满足一定拓扑结构; 3) 所有的可行解中 $f(H)$ 值最大。

2.2.1.1 持续社区搜索模式

为了体现时序属性图上的社区的动态性, Li 等^[21] 提出了极大 (θ, k) -persistent-core 间隔, 并在此基础上定义了 (θ, τ) -persistent-core 结构。

定义 26 (θ, τ) -persistent-core 结构

给定一个时序属性图 G 和参数 θ, τ, k , 一个

(θ, τ) -persistent-core 是一个 G 的诱导时序子图 $C(V_C, E_C)$, 满足: 1) $F(\theta, k, C) \geq \tau$ ^[21]; 2) 不存在诱导时间子图 H 包含 C 且满足性质 1)。

模式 13 持续社区搜索 (persistent community search, PC) 模式

给定一个时序属性图 G_T 和参数 θ, τ, k 。搜索出 G 中极大 (θ, τ) -persistent-core。

Li 等^[21] 给出一种剪枝搜索算法, 通过将时序属性图的整个时间跨度分解成多个子区间, 每个子区间都有很多可剪枝的节点。

2.2.1.2 最小化低效率值社区搜索模式^[22]

模式 14 最小化低效率值时序社区搜索模式

给定一个时序属性图 $G_T=(V, E, \Sigma, L)$ 、一个参数 $\alpha \in [0, 1]$ 、一个查询节点集 $Q \subseteq V$, 旨在找到一个最小低效率时序子图。

该模式的核心概念为时序网络低效率值^[23]。Ruchansky 等^[22] 通过图形变换、时序连接、最小化低效率时序子图^[23] 共 3 个步骤解决了该问题。

2.2.2 位置属性图社区搜索模式

位置属性图即将地理位置信息作为节点属性的图, 如地理社交网络等。

定义 27 位置属性图社区搜索模式

给定一个位置属性图 $G_C=(V, E, \Sigma, L)$ 、一个查询节点集 $Q \subseteq V$ 、一个属性评分函数 f , 需要找到一个 G 的诱导子图 $H(V_H, E_H, \Sigma_H, L_H)$, 满足如下性质: 1) V_H 包含 Q ; 2) H 是连通的, 且满足一定拓扑结构; 3) 所有可行解中 $f(H)$ 值最大。

2.2.2.1 限定子图位置的社区搜索模式

模式 15 限定子图地理位置的社区搜索模式
给定一个位置属性图 $G_C=(V, E, \Sigma, L)$ 、一个查

询节点 $q \in V$ 、一个空间约束 A , 返回一个子图 $G_q \subseteq G_C$, 满足以下条件: 1) 连通性, G_q 中包含 q 且是连通的; 2) 结构内聚性, G_q 满足一定的拓扑结构; 3) 空间内聚性, G_q 满足空间内聚约束 A 。

对于不同模式, 性质 2) 结构内聚性和性质 3) 空间内聚性可能是不同的。SAC 搜索模式、RB k -core 搜索模式、地理社交社区查询 (geo-social group queries, GSGQs) 模式都属于限定子图地理位置的社区搜索模式, 相同之处在于使用同样的拓扑结构 k -core 作为结构内聚性约束; 不同之处在于所用的约束空间内聚性的条件不同。

1) SAC 搜索模式所采用的空间约束 A 要求返回社区的最小覆盖圆 (minimal cover circle, MCC)^[24] 的半径最小。Guo 等^[24] 根据密克定理开发出的 2 种近似算法一定程度上提高了搜索效率。

2) RB k -core 搜索模式所采用的空间约束 A 也是 MCC 但其半径需要小于一个给定值 r , 且需要再加入一条极大性约束: 不存在其他的子图 G'_q 满足上述性质且有 $G_q \subseteq G'_q$ 。Wang 等^[25] 为该模式提供了 TriV、BinV 和 RotC 共 3 种算法。

3) GSGQs 模式中, Zhu 等^[26] 所使用的空间约束 A 包括 3 种: 1) A 是用于包含 G_q 的空间矩形; 2) A 是以 q 为中心的圆; 3) A 是满足约束 2) 的圆, 且 G_q 恰好包含 $k+1$ 个节点。Zhu 等^[26] 提出了 social-aware r-trees 索引, 基于该索引开发了高效的算法来解决具有不同空间约束的 GSGQs。

对限定子图位置的社区搜索所用到的 3 个模式的解决算法的时间复杂度总结如表 2。表 2 中: n 为节点数量, m 为边数量, ϵ_F 为给定参数, k 为最小度。

表 2 限定子图位置的 3 种社区搜索模式对比

Table 2 Comparison of three community search models that limit the location of subgraphs

模式	空间限制	算法	时间复杂度
SAC ^[24,27-28]	MCC	Exact	$O(m \cdot n^3)$
		AppInc	$O(mn)$
		AppFast	$O\left(m \cdot \min\left\{n, \log \frac{1}{\epsilon_F}\right\}\right)_{\epsilon_F > 0}$ $O(mn)_{\epsilon_F = 0}$
		AppAcc	$O\left(\frac{m}{\epsilon_F} \cdot \min\left\{n, \log \frac{1}{\epsilon_F}\right\}\right)$
		Exact+	$O\left(\frac{m}{\epsilon_F} \cdot \min\left\{n, \log \frac{1}{\epsilon_F}\right\} + m F_1 ^2\right)$
RB k -core ^[25]	半径固定的圆	TriV	$O(n^3 \cdot (m + n))$
		BinV	$O(n^2 \cdot (m + n))$
		RotC	$O(n^2 \cdot (\log n + m'))$

续表 2

模式	空间限制	算法	时间复杂度
GSGQs ^[26]	矩形, 中心固定的圆	CBR	$O(m \cdot \log n)$
		GSGQ _{range}	$O(m + n)$
		GSGQ _{rkNN}	$O(n \cdot (m + n))$
		GSGQ _{kNN}	$O(C_k^{n-1} (m + n))$

2.2.2.2 共位社区搜索模式

共位社区是一个具有高空间内聚度的连通图。

定义 28 共位社区^[27]

给定一个位置属性图 $G_C=(V, E, \Sigma, L)$, 共位社区是一个子图 $J \subseteq G_C$, 且满足: 1) 连通性, J 是连通的; 2) 结构内聚性, 子图 J 满足一定的内聚性结构; 3) 空间内聚性, J 中所有节点在空间上相临近。

模式 16 最佳共位社区^[28](best co-located community, BCC)搜索模式

给定一个位置属性图 $G_C=(V, E, \Sigma, L)$ 、一个关键字属性集 T 、一个正整数 k 、一个直径 D , BCC 搜索模式旨在找到一个子图 G_k 满足如下性质: 1) 结构内聚性为 k -core; 2) 节点的位置可以被直径为 D 的圆包围; 3) 质量 $Q(G_k)$ 在满足前面 2 个条件中是最大的。

BCC 搜索模式包含 2 种基线算法 BSL1、BSL2 以及基于索引 AC-tree 算法共 3 种算法。

模式 17 最具结构内聚性的共位社区 (most cohesive co-located community, MC³) 搜索模式^[29]

给定一个位置属性图 $G_C=(V, E, \Sigma, L)$ 、一个关键字属性集 T 、直径 D , MC³ 搜索模式旨在找到一个子图 G_k 满足: 1) G_k 是一个 k -core; 2) G_k 中节点位置可以被一个直径为 D 的圆包围; 3) 在所有满足上述 2 个条件的可行解中, G_k 是一个具有最大 k 值的 k -core。

MC³ 搜索模式包含一种基线算法以及一种基

于四叉树索引 DkQ-tree 的搜索算法^[29]。

模式 18 (k, d)-MCCs 搜索模式

给定一个位置属性图 $G_C=(V, E, \Sigma, L)$ 、一个正整数 k 和参数 d , 返回一个共位社区 $H(V_H, E_H)$, 满足: 1) H 是一个 k -truss; 2) H 中任意 2 点之间的空间距离不大于 d ; 3) 不存在其他的共位社区 H' 满足上述条件且有 $|V(H')| > |V(H)|$ 。

(k, d)-MCCs 搜索模式包含一个基线算法、一个近似算法^[29] 以及一个基于四叉树的索引算法。

2.2.2.3 位置属性搜索模式关系

由定义 14~17 可得, 本节介绍的几种位置属性图中的搜索模式 (k, d)-MCCs 搜索模式从属于 RB k -core、SAC、BCC、GSGQs 模式, 其中 RB k -core、SAC、BCC、GSGQs 模式相互等价。其关系表示如图 2。

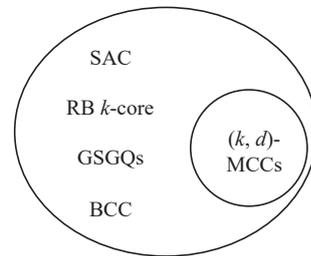


图 2 位置属性 CS 模式关系

Fig. 2 Relationship between some location_based CS schema

共位社区搜索的 3 种典型模式的异同点总结如表 3 所示。

表 3 共位社区搜索的 3 种典型模式对比

Table 3 Comparison of three typical co-located community search models

模式	空间约束	属性评分函数	内聚结构	最大化约束
BCC ^[28]	空间圆	√	k -core	质量 Q
MC ³ [28-29]	空间圆	√	k -core	k 值最大
(k, d)-MCCs ^[29]	给定最大节点距离	×	k -truss	$ V $ 最大

2.3 关键字、权值属性图社区搜索模式

具有关键字、权值属性的属性图分别被称为关键字属性图和权值属性图, 许多社区搜索问题的提出建立在这 2 种属性图中。

2.3.1 关键字属性图社区搜索模式

关键字属性图, 即将关键字作为节点属性的

图, 研究关键字属性图中的社区搜索模式可以直观地找到在复杂网络中具有共性的社区。

定义 29 关键字属性图社区搜索模式

给定一个关键字属性图 $G_K=(V, E, \Sigma, L)$ 、一个查询节点集 $Q \subseteq V$ 、一个属性评分函数 f , 需要找到一个 G 的诱导子图 $H(V_H, E_H, \Sigma_H, L_H)$, 满足如

下性质: 1) V_H 包含 Q ; 2) H 是连通的, 且满足一定拓扑结构; 3) 所有的可行解中 $f(H)$ 值最大。

2.3.1.1 传统的关键字属性图社区搜索模式

传统的关键字属性图上的社区搜索模式可以概括为给定一个关键字属性图、一个查询节点集、一个查询属性集和一个与关键字相关的属性评分函数, 返回一个或多个满足一定拓扑结构的子图。

模式 19 属性社区查询(attribute community query, ACQ)模式^[30]

给定一个关键字属性图 $G_K=(V, E, \Sigma, L)$ 、一个正整数 k 、一个查询节点 $q \in V$ 以及一个查询属性集 $S \subseteq W(q)$, ACQ 模式需要返回一个元素为 G_K 的子图 G_q 所构成的集合 S 。 $\forall G_q \in H$, 需要满足: 1) 连通性, G_q 是连通的且包含查询节点 q ; 2) 结构内聚性, G_q 是一个 k -core; 3) 关键字内聚性, S 与 G_q 中所有元素共有的所有关键字属性的集合 $L(G_q, S)$ 中的元素是最多的。

最简单处理 ACQ 模式的算法为枚举法, Fang 等^[30] 还提出了一个基于索引 CL-tree 的搜索算法, 其空间和时间复杂度几乎达到了初始图的线性水平。

模式 20 属性驱动的 truss 社区(attribute-driven truss community, ATC)搜索模式^[31]

给定一个关键字属性图 $G_K=(V, E, \Sigma, L)$ 、一个包含查询节点 V_q 和目标关键字 W_q 的查询集 $Q=(V_q, W_q)$ 以及 2 个常数 k 和 d , 返回一个社区 H 满足: 1) H 是包含 V_q 的 (k, d) -truss; 2) 满足上一条性质的子图中, H 的属性得分 $f(H, W_q)$ 最大^[31]。

ATC 搜索模式包含一种全局搜索算法以及一种基于索引的算法, Huang 等^[31] 提出了另一种基于 ATindex 的搜索算法, 可以快速识别出一个好的候选答案。

2.3.1.2 极小覆盖 r -clique 社区搜索模式

定义 30 覆盖 r -clique

一个覆盖 r -clique C 是给定图 G 中具有所有查询关键字的节点的集合, 其中任意一对关键字节点之间的最短距离不大于 r 。

定义 31 极小覆盖 r -clique(minimal covered r -clique, MCCr)^[32-33]

当满足不存在其他的 r -clique $C' \subseteq C$, 则称 C 为 G 的极小覆盖 r -clique。

极小覆盖 r -clique 社区搜索模式^[32] 目标为在给定图 G 中找到具有目标关键字的极小覆盖 r -clique 社区。

该模式包含 Bron-Kerbosch 关键字搜索(Bron-

Kerbosch-based keyword search, BKS)算法, 该算法添加预处理, 并构建递归树, 来提高广度优先算法(breadth first search, BFS)效率; 以及一种改进了递归树后提出的 BKSR 算法, 进一步提高了搜索效率。提出 BKSRM 算法, 当迭代次数为 T 时, 时间复杂度减小到了 $\frac{O(|C_{\max}|^{t+1})}{T}$ 。

2.3.1.3 属性多样化的社区搜索模式

基于关键字属性的社区搜索模式往往以寻找节点关键字相似的社区为目标, 另有属性多样化社区搜索模式目标在于找到具有尽量多不同属性的社区。如在一场高校举办的联谊会上, 举办者可能希望更多所在不同学院的学生相互联谊。

模式 21 属性多样化社区搜索模式(attribute diversified community search, ADCS)^[33]

给定关键字属性图 $G_K=(V, E, \Sigma, L)$, 返回子图 H 具有性质: 1) H 满足 k -core 结构; 2) H 的平均边多样性在所有满足条件 1) 中的社区中是最大的。

ADCS 问题目前有基线枚举算法、advADC 共 2 种解决方法, 与前者相比, 添加了额外的循环, 以更小的成本修建 k -core, 使得迭代的次数大大减小。

大部分属性图社区搜索模式都会对不同属性构建不一样的属性评分函数。属性评分函数在社区搜索问题中应用极广, 例如 2.3.1.1 节中 $f(H, W_q)$ ^[31] 以及 $f=|L(G_q, S)|$, 它可以将搜索需求公式化, 以函数的计算结果反映返回社区与需求的契合程度。但是, 大多数社区搜索问题的属性评分函数难以全方位考虑到所有影响因素。Wang 等^[34] 在评分函数 $f(H, W_q)$ 的基础上提出了新的属性评分函数 $\text{AttriScore}(H, W_q) = |W_q| \sqrt{\prod_{w \in W_q} \text{score}(H, w)}$, 它满足了 Wang 等^[34] 提出的构造属性评分函数的 4 个准则:

- 1) 与查询属性相关的节点越多的社区得分更高;
- 2) 覆盖更多查询属性的节点所在社区得分更高;
- 3) 如果有一个节点只覆盖很少量的属性, 其所在的社区的得分函数应该更低;
- 4) 规模更大的社区与规模较小的社区, 前者的得分应该更高。

Liu 等^[35] 为了将单一属性下的社区搜索延伸到多属性, 提出一种属性评分函数 $\max_{u, v \in V_G} \text{score}(u, v)$ ^[35], 旨在衡量包含多属性图中的社区搜索所返回的社区的优劣。

2.3.1.4 关键字属性图社区搜索模式关系

2.3.1.1 节定义的几种模式下的问题中, ATC 搜索问题与 ADCS 及 ACQ 问题的输入初始图属性都是关键字属性图, 输出社区的拓扑结构分别

为 k -truss、 k -core 及 k -clique。则根据定义 14~17, ACQ 模式与 ADCS 模式相互等价且 ATC 模式从属于 ACQ 与 ADCS 模式。其关系表示如图 3, 读者在选择相应搜索模式时, 应考虑其从属关系, 选择详细或简略的搜索模式。

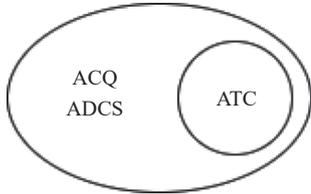


图 3 关键字属性 CS 模式关系

Fig. 3 Relationship between some keyword based CS schema

2.3.2 权值属性社区搜索模式

权值属性图即以权值为属性的图结构。

定义 32 权值属性图社区搜索模式

给定一个权值属性图 $G_W=(V, E, \Sigma, L)$ 、一个查询节点集 $Q \subseteq V$ 、一个属性评分函数 f , 需要找到一个 G 的诱导子图 $H(V_H, E_H, \Sigma_H, L_H)$, 满足如下性质: 1) $Q \subseteq V_H$; 2) H 是连通的, 且满足一定拓扑结构; 3) 所有的可行解中 $f(H)$ 值最大。

这里的权值可以为节点的权值或是边的权值。下文将以节点权值为属性的图称为点权属性图, 边的权值为属性的图称为边权属性图。权值在不同的模式中也可以称为重要性、影响力等。

2.3.2.1 点权属性社区搜索模式

点权属性图是一个无向图 $G=(V, E)$, 其中每个节点 $v \in V$, 都对应一个权值 w_v , 代表着节点 v 的重要性。假设将所有权重组成向量 $W=(w_1, w_2, \dots, w_n)$, 且满足降(增)序排列, 需要满足: 对于任意 2 个节点 v_i 和 v_j , 如果 $i \neq j$, 那么 $w_i \neq w_j$ ^[36]。

每个点至多对应一个权值的属性图称为一维点权属性图, 对应一个以上权值的属性图称为多维点权属性图^[37], 多维点权属性图是一个多值图 $G(V, E, X)$, 其中 V, E 分别代表节点集和边集, X 是 d 维向量组成的一个集合。在多值图 $G(V, E, X)$ 中, 每个节点 $v \in V$ 与 d 维向量 $X_v=(x_1^v, x_2^v, \dots, x_d^v)$ 相对应。

点权属性社区搜索模式可细分为以下 3 类模式:

1) 一维点权属性图中的社区搜索模式

一维点权属性图 G 中, 图权值 $W(G)$ 即图 G 中所有节点权值的最小值, 基于图权值定义了 k -influential 社区、非包含 k -influential 社区^[38-39] 2 类社区。

定义 33 k -influential 社区^[36]

给定一个点权属性图 $G_W=(V, E, \Sigma, L)$ 、一个整数 k , 则 k -influential 社区 $H^k=(V_H^k, E_H^k)$ 为 G_W 的一个诱导子图, 满足: 1) 连通性, H^k 是连通的; 2) 内聚性, H^k 中每个节点的度数至少是 k ; 3) 极大性, 没有其他诱导子图 \bar{H} 在满足性质 1)、性质 2) 和 \bar{H} 包含 H^k 的前提下, 有 $f(H^k) = f(\bar{H})$, 这里的 $f(H) = \min_{u \in V_H} \{w_u\}$ 。

模式 22 top- k -influential 社区搜索模式

给定一个点权属性图 $G_W=(V, E, \Sigma, L)$ 、一个整数 k , 找到 f 的大小为 top- k 的 k -influential 社区。

定义 34 非包含 k -influential 社区

非包含 k -influential 社区 H^k 是一个 k -influential 社区且 H^k 不能包含一个 \bar{H}^k 满足 $f(\bar{H}^k) = f(H^k)$ 。

模式 23 top- k 非包含 k -influential 社区搜索模式

给定一个点权属性图 $G_W=(V, E, \Sigma, L)$ 、一个整数 k , 找到权值最高的 top- k 个非包含 k -influential 社区。

Li 等^[36] 给出了一种在线搜索算法, Chen 等^[40] 提出一种反向搜索算法, 将没有希望成为结果社区的搜索分支提前终止。Li 等^[36] 还给出了一种基于权值社区保存索引 (influential-community preserved index, ICP-index) 的搜索算法, 一定程度上提高了搜索效率。

2) 多维点权属性图上的社区搜索模式

多维点权属性图中, 所有的节点 $v \in V, x_i^v$ 形成了一个严格的顺序。即对于任意 $u \neq v, x_i^v \neq x_i^u$ 。若多维点权属性图 G 的所有诱导子图 $H, H_1, H_2, \dots, H_n, H_i < H$ 对 $\forall i \in [1, n]$ 成立, 称 H 为 skyline 社区^[37]。

模式 24 Skyline 社区搜索模式^[37]

给定一个多值图 $G(V, E, X)$ 和一个整数 k , 该模式旨在找到一个 Skyline 社区, 它是 G 的一个诱导子图 $H=(V_H, E_H, X_H)$, 满足: ① H 是一个连通的 k -core; ② 不存在 G 的子图 H' 是连通的 k -core 且有 $H < H'$; ③ 极大性: 对任意 $1 \leq i \leq d$, 不存在 G 的子图 H' 满足上述条件且有 $f_i(H) = f_i(H'), f(H) = \min_{v \in V_H} \{x_i^v\}$ 。

对于该模式, Li 等^[37] 开发了名为 Skyline Comm2D 的高效算法, 它的时间复杂度是 $O(s(m+n))$, 这里的 s 指的是所找到的 skyline 社区的数量。为了能够更好地处理高维的情况, Li 等提出了一种空间划分算法来有效地寻找 Skyline 社区。

3) 点权属性社区搜索模式关系

由定义 14~17 可得, 本小节介绍的几种点权属性图中的搜索模式中, Skyline CS 模式与 TIC

搜索模式具有相交的关系, 且二者与 PC、MTIC 搜索模式之间不相关, 其关系表示如图 4。

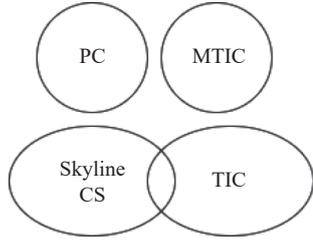


图 4 点权属性 CS 模式关系

Fig. 4 Relationship between some node weighted based CS schema

2.3.2.2 边权属性社区搜索模式

根据定义 6, 边权属性图是一个权值属性图 $G_w = (V, E, \Sigma, L)$, 其中每个节点 $e \in E$, 都对应一个权值 w_e , 可以表示边 e 的权重、构成该边的节点之间的紧密程度、传播概率等。本节将边权属性图分为了常规边权属性图、二部边权属性图和不确定图。

1) 常规边权属性图中的社区搜索模式

为了适合在常规边权属性图上社区搜索, 以下拓扑结构被提出: 加权 truss^[41]、 pk -clique^[42]、 kr -clique^[2]、 pkd -truss^[43]、紧密 core 组^[44]。

定义 35 加权 truss 社区 (weighted truss community, WTC)

给定一个边权属性图 $G_w = (V, E, \Sigma, L)$ 和一个正整数 k , WTC 是 G_w 的一个诱导子图 H , 具有以下性质: ① $\forall e_1, e_2 \in E(H)$, 满足 e_1, e_2 在 H 中是三角连通的; ② H 是一个 k -truss; ③ H 是满足上述性质的极大子图, 即在 G_w 中不存在其他 $H \subset \bar{H}$, 且这 2 个子图的权值相等。

模式 25 WTC 搜索模式

给定一个边权属性图 $G_w = (V, E, \Sigma, L)$ 、一个正整数 k , 找出权值大小为 top- k 的 WTC。

为了提高搜索效率, Zheng 等^[41] 设计了一个索引结构 KEP-index, 基于 KEP-index, WTC search 可以在线性时间内完成。

定义 36 pk -clique

给定图 G 、参数 p 和 k , pk -clique $C_{pk}(V_{pk}, E_{pk})$ 是 G 的一个诱导子图且满足以下性质: ① C_{pk} 是一个完全图; ② C_{pk} 中任意 2 个节点可以通过一条 MIP 相连, 其路径权值大于 p , 且 C_{pk} 中至少要有 k 个节点; ③ C_{pk} 是 G 的所有满足上述条件的 pk -clique 中节点最多的。

模式 26 Top- k 影响力社区搜索模式^[42]

给定一个边权属性图 $G_w = (V, E, \Sigma, L)$ 、一个正整数 k , 找到对查询节点 s 影响最大的 top- k 个 pk -clique。

定义 37 kr -clique

给定图 G 和 2 个整数 k, r , kr -clique H 满足如下条件: ① H 是 G 的诱导子图; ② H 至少有 k 个节点; ③ H 中任一对节点可以在 r 次跳跃内到达彼此。

模式 27 最具影响力的社区搜索模式 (most influential community search, MICS)

给定一个边权属性图 $G_w = (V, E, \Sigma, L)$ 、整数 k, r 以及一个常数 t , 返回一个 kr -clique H , 在所有满足要求的 kr -clique 中 H 的得分 $\text{score}(H)$ 最高。

该模式包括一种基线算法, Li 等^[2] 提出了一种树形索引 C-Tree, 基于 C-Tree 并使用 DFS 的算法 SO 和 SO+, 可以显著地提高搜索最大 kr -clique 的速度。鉴于 DFS 的特性, Li 等^[2] 又开发出了一种算法, 能够以更稳定的速度搜索到 kr -clique。

定义 38 pkd -truss 社区

该社区是一个 (k, d) -truss^[43], 且满足任意节点到给定节点之间的最大路径影响值不小于 p 。

Xie 等^[43] 以 pkd -truss 结构为基础提出了 pkd -truss 社区搜索模式。

模式 28 pkd -truss 社区搜索模式

给定一个边权属性图 $G_w = (V, E, \Sigma, L)$ 、一个查询节点集 (V_q, W_q) 、参数 p, k, d , 该模式目标是找到一个 pkd -truss 社区 H , 其得分最高。

定义 39 紧密 core 组

给定边权属性图 G 、一个查询节点集、一个参数 k , 一个紧密 core 组社区是 G 的子图 H 满足下面条件: ① H 包含查询节点集中的所有元素; ② H 是一个连通的 k -core; ③ H 是满足上述 2 条件中图权值最小的, 这里的图权值定义为图中所有边权之和。

模式 29 紧密 core 组搜索 (intimate-core group search, LEKS) 模式^[44]

给定边权属性图 $G_w = (V, E, \Sigma, L)$ 、一个查询节点集 Q 、一个参数 k , 目标在于找到 G 的子图 H 满足: ① H 包含查询节点集中的所有元素; ② H 是一个连通的 k -core; ③ H 是满足上述 2 条件中图权值最小的。对于 LEKS 模式, Sun 等^[44] 给出了 2 种在线搜索算法 ICG-S 和 ICG-M, 并提出了一种基于 k -core 索引的搜索算法提高了搜索效率。

2) 二部边权属性图中的社区搜索模式

由于二部图的特殊性, 在二部边权属性图上基于传统拓扑结构的社区搜索模式相对较少。有很多由传统拓扑结构衍生出来的新结构被设计出来应用在二部图中, 如 (α, β) -core^[45]、bitruss^[46]、biclique^[47]。

将二部边权属性图 G 的 2 个互不相交的点集

分别记为上层 (upper layer) $U(G)$ 和 下层 (lower layer) $L(G)$, 则首先给出 (α, β) -core 的定义:

定义 40 (α, β) -core

图 G 的 (α, β) -core $R_{\alpha, \beta}$ 满足 $R_{\alpha, \beta}$: ① $\deg_{U(G)}(u) \geq \alpha$, $\deg_{L(G)}(u) \geq \beta$; ② 不存在其他的子图 H 满足性质 ① 且有 $R_{\alpha, \beta} \subseteq H$ 。

模式 30 重要 (α, β) 社区 (significant (α, β) -community, SC) 搜索模式^[45]

给定一个二部边权属性图 G , 找到一个具有最大边权的 (α, β) -core 子图。

针对该模式, Liu 等^[45] 先提出了基于索引 $I_{b_s}^\alpha$ 的搜索算法和基于 $I_{b_s}^\beta$ 的搜索算法, 又提出了一种基于退化有界索引 I_δ 的搜索算法, 有效地解决了当二部图中部分节点度数过高导致的搜索效率过低的问题。

3) 不确定图上的属性图社区搜索模式

大部分的社区搜索问题都在确定图上被提出, 即假定各个节点之间的关系是固定的。这种假设忽略了节点之间的不确定性。因此, 在不确定图^[48-50] 中研究社区搜索模式是很有必要的。

定义 41 (k, η) -influential 社区

给定一个不确定图 G 、参数 k, η , (k, η) -influential 社区是 G 的最大子图, H 满足: ① H 是连通的; ② $\forall v \in H, v$ 至少有 k 个邻居的概率不小于 η ; ③ 不存在其他 G 的诱导子图 H' 满足上面 2 个条件且有 $f(H') = f(H)$, 这里 $f(H) = \min_{v \in V_H} \{w(v)\}$ ^[49]。

模式 31 (k, η) -influential 社区搜索模式

给定一个不确定图 G , 一个查询节点 q , 参数 k, η , 找到一个 (k, η) -influential 社区。

为了解决 (k, η) -influential 社区搜索问题, Luo 等^[49] 提出了在线搜索算法、基于 ICU-index 的搜索以及基于 FICU-index 的算法共 3 种搜索算法。

定义 42 (k, θ) -community

给定一个不确定图 G , 查询节点 q , 2 个参数 k, θ 。其中 $k \geq 2$, 且 k 为正整数, $0 < \theta < 1$ 。 (k, θ) -community H 是 G 的一个连通子图, 且满足: ① $q \in H$; ② $\forall v \in H, v$ 有至少有 k 个邻居; ③ $\forall e \in H, e$ 存在的概率不小于 θ ; 4) 在所有满足上面条件的 H_i 中, H 包含最多的顶点。

模式 32 (k, θ) -community 搜索模式

给定一个不确定图 G , 查询节点 q , 2 个参数 k, θ , 找到一个 (k, θ) -community^[50]。

为了解决 (k, θ) -community 搜索问题, Miao 等^[50] 首先提出了 GSP、NCP 共 2 种剪枝策略, 并提出 CD-index 以高效完成剪枝过程, 一定程度上缩小了候选集的规模。进而提出了一种惰性分层

抽样算法, 在保证精度的前提下加快了搜索速度。

4) 边权属性图社区搜索模式关系

由定义 14~17 可得, 本小节介绍的几种边权属性图中的搜索模式中: MICS 模式从属于 (k, d) -MCCs 和 WTC 模式; (k, d) -MCCs 和 WTC 模式相互等价且从属于 LEKS、SC、 (k, η) -influential CS 搜索模式, 且这 3 种模式相互等价。其关系表示如图 5 所示。

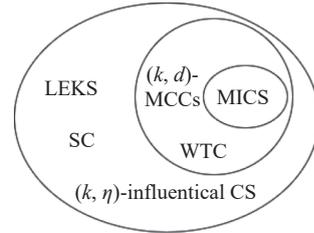


图 5 边权属性 CS 模式关系

Fig. 5 Relationship between edge weighted-based CS schema

权值 (点权、边权) 作为不仅是图模式中最常见的属性, 在各类研究^[51-54] 中都经常用到。社区搜索在权值属性图中的研究相较于其他属性也更多。如表 4 对权值属性图中的社区搜索的一些经典模式以及其算法时间复杂度进行了对比。但是在解决实际问题中, 不同模式在解决不同社区搜索问题时有着独特的优势, 并不一定以时间复杂度作为最主要的参考依据。表 4 中, n 为顶点数量, m 为边数量, s 为 skyline 社区的数量, ρ 为 AWG 的荫度, r 为迭代次数。

表 4 几种权值属性图社区搜索模式对比

Table 4 Comparison of several community search models in weighted graphs

模式	算法	时间复杂度
TIC	Union-Find	$O(n + m \cdot \log n)$
	CountIC	$O(m + n)$
	EnumtIC	$O(m + n)$
TnIC	NC1	$O(m)$
	NC2	$O(m)$
Skyline CS	Comm2D	$O(s(m + n))$
WTC	基于KEP-index	$O(m^{1.5} + k_{\max} \cdot n)$
	MCP	$O(3^{n/3})$
pkd-truss CS	IACSM	$O(n^2 + \rho m + rn \cdot \text{AttriList}_q)$
SC	Basic: $I_{b_s}^\alpha$	$O(\partial_{\max} \cdot m)$
	Basic: $I_{b_s}^\beta$	$O(\beta_{\max} \cdot m)$

3 属性图社区搜索模式分析与检索

由于存在过多的属性社区搜索模式, 研究者

难以根据需求快速选择其中最合适的。本节分析各种模式的输入属性图属性、输出社区的拓扑结构、是否用到属性评分函数以及现存的算法种类, 得出表 5。学者们可根据需求, 依据表 5 快速得到适合的属性社区搜索模式。

表 5 模式对比
Table 5 Schemas comparison

模式	属性函数	输入属性	算法			输出
			局部	全局	基于索引	
19	√	关键字	×	×	√	<i>k</i> -core
20	√	关键字	√	√	√	<i>k</i> -truss
21	√	关键字	√	×	√	<i>k</i> -core
15	√	位置	√	√	√	<i>k</i> -core
16	√	位置	√	×	√	<i>k</i> -core
17	×	位置	×	√	√	<i>k</i> -core
18	×	位置	×	√	√	<i>k</i> -truss
13	×	时序	×	√	×	<i>k</i> -core
14	√	时序	×	√	×	<i>k</i> -clique
22	√	权值	×	√	√	<i>k</i> -core
24	√	权值	×	√	×	<i>k</i> -core
25	×	权值	×	√	√	<i>k</i> -truss
26	×	权值	×	√	√	<i>k</i> -clique
27	√	权值	×	√	√	<i>k</i> -clique
28	√	权值	√	×	×	<i>k</i> -truss
29	×	权值	×	√	√	<i>k</i> -core
30	×	权值	×	√	×	<i>k</i> -core
31	√	权值	×	√	√	<i>k</i> -core

3.1 属性图社区搜索模式特性分析

本节将非空属性图中的社区搜索进行对比分析。表 5 描述了每一种模式的输入属性图的类别、输出的拓扑结构类别、使用属性函数与否、该类模式中的现有算法。如在 ACQ 模式中, 其输入是一张关键字属性图, 该模式包括了一种基于索引的算法, 最终输出一张满足拓扑结构为 *k*-core 的子图, 并使用了属性评分函数来约束子图的优劣。

3.2 属性图社区搜索模式检索

当明确了已知问题的前置条件后, 本文可以帮助研究者在众多的社区搜索问题中准确定位到需要的解决方法, 具体实现方法如下例。

图 6(a) 数据来源于 DBLP 网络^[55], 共计 9824681 条数据, 本文为了清晰地呈现数据, 截取了 2000 条数据构建了权值属性图。其中节点表示学者, 标签表示学者的姓名(为了展现清晰, 标签未标注), 边表示相邻学者之间有过合著论文的关系,

边权表示合著论文的篇数。将图 6(a) 为初始图输入, 给定查询节点 *q* 代表作者如 Francesco Palmirei, 要求尽快搜索到结构紧密且与查询节点所代表的作者之间产生合著关系尽可能多的社区, 若要求返回子图为 *k*-core 结构。

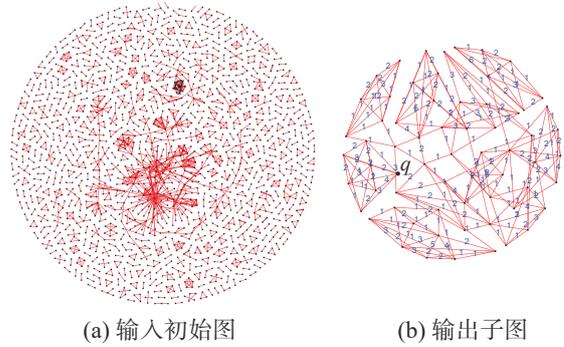


图 6 社区搜索示例

Fig. 6 Example of community search

1) 根据要求, 在表 5 中选定条件分别为“权值”、“*k*-core”, 则返回满足条件的社区搜索模式为 LEKS 模式、SC 搜索模式或 (*k*, η)-influential CS 模式。

2) 为了追求更快的搜索速度, 排除不具有基于索引算法的 SC 搜索模式。再定位到模式 29 及模式 30, 由于 (*k*, η)-influential CS 模式更适用于二部图中的社区搜索, 故选定 LEKS 模式解决该问题。

3) 选定 LEKS 模式并选用该模式的 ICG-S 算法, 给定参数 *k*=5, 最终返回社区如图 6(b), 社区满足包含查询结点 Francesco Palmirei 的 5-core 社区中图权值最大的。

LEKS 模式中要求返回社区的图权值最小, 本例中更改 LEKS 模式中条件 3) 为返回社区 *H* 是满足上述 2 条件中图权值最大的。在 DBLP 网络中进行社区搜索可以迅速找到与目标作者合著最多的作者所在的社区, 帮助学者迅速、准确地找到相关方向论文。

4 结束语

综上所述, 随着图查询研究的逐步深入, 社区搜索作为图查询研究中极其重要的一环, 已被广泛应用于诸多实际应用中, 具有重要的理论和实践意义。本文首先从社区搜索问题的要素入手, 定义了属性图中各种社区搜索模式及其之间的关系; 其次分析了各种属性图社区搜索模式并对相关算法进行对比分析, 进一步提出了属性图社区搜索模式的检索方式, 能够以简单、在线的方式找到满足用户需求的社区搜索模式。

本文在前人总结的社区搜索问题的基础上提

出了属性图中的社区搜索模式,并提出了两层分类体系,以帮助研究者更准确地定位所需的算法。这种全新的模式为社区搜索领域的研究者提供了一个更具体和系统的框架,通过提供这种模式总结和分类方法,为研究者提供了一个全面的概念视角,并将他们从烦杂的综述中解放出来,能够更快速地找到和应用适合需求的算法,提高了社区搜索领域的研究效率和准确性。

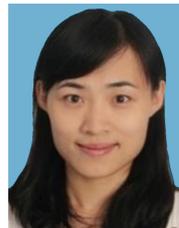
参考文献:

- [1] LI Xiaoli, WU Min, KWOH Chee-keong, et al. Computational approaches for detecting protein complexes from protein interaction networks: a survey[J]. *BMC genomics*, 2010, 11(1): 1–19.
- [2] LI Jiabin, WANG Xinjue, DENG Ke, et al. Most influential community search over large social networks[C]//2017 IEEE 33rd International Conference on Data Engineering. San Diego: IEEE, 2017: 871–882.
- [3] CAI Liwei, WANG Yang. KBGAN: adversarial learning for knowledge graph embeddings[C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. New Orleans: NAACL, 2018: 1470–1480.
- [4] SOZIO Mauro, GIONIS Aristides. The community search problem and how to plan a successful cocktail party[C]//Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data mining. Washington DC: KDD, 2010: 939–948.
- [5] FANG Yixiang, HUANG Xin, QIN Lin, et al. A survey of community search over big graphs[J]. *The VLDB journal*, 2020, 29(1): 353–392.
- [6] BARBIERI N, BONCHI F, GALIMBERTI E, et al. Efficient and effective community search[J]. *Data mining and knowledge discovery*, 2015, 29(5): 1406–1433.
- [7] CUI Wanyun, XIAO Yanghua, Wang Haixun, et al. Local search of communities in large graphs[C]//Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data. Snowbird: MOD, 2014: 991–1002.
- [8] ZHU Junchao, WANG Chaokun. Approaches to community search under complex conditions[J]. *Journal of software*, 2019, 30(3): 552–572.
- [9] FANG Yixiang, WANG Zhihong, CHENG Reynold, et al. Effective and efficient community search over large directed graphs[J]. *IEEE transactions on knowledge and data engineering*, 2019, 31(11): 2093–2107.
- [10] LIU Qing, ZHAO Minjun, HUANG Xin, et al. Truss-based community search over large directed graphs[C]//Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data. New York: MOD, 2020: 2183–2197.
- [11] LI Zhenjun, LU Yunting, ZHANG Weipeng, et al. Discovering hierarchical subgraphs of k-core-truss[J]. *Data science and engineering*, 2018, 3(2): 136–149.
- [12] HUANG Xin, CHENG Hong, QIN Lin, et al. Querying k-truss community in large and dynamic graphs[C]//Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data. Snowbird: SIGMOD, 2014: 1311–1322.
- [13] AKBAS Esra, ZHAO Peixiang. Truss-based community search: a truss-equivalence based indexing approach[J]. *Proceedings of the VLDB endowment*, 2017, 10(11): 1298–1309.
- [14] HUANG Xin, LAKSHMANAN L V S, YU J X, et al. . Approximate closest community search in networks [EB/OL]. (2015-03-22)[2023-03-25]. <https://arxiv.org/abs/1505.05956.pdf>.
- [15] YUAN Long, QIN Lu, LIN Xuemin, et al. Diversified top-k clique search[J]. *The VLDB journal*, 2016, 25(2): 171–196.
- [16] WU Jun, LI Chumin, JIANG Lu, et al. Local search for diversified Top-k clique search problem[J]. *Computers & operations research*, 2020, 116: 104867.
- [17] PALLA G, DERÉNYI I, FARKAS I, et al. Uncovering the overlapping community structure of complex networks in nature and society[J]. *Nature*, 2005, 435(7043): 814–818.
- [18] YUAN Long, QIN Lu, ZHANG Wenjie, et al. Index-based densest clique percolation community search in networks[J]. *IEEE transactions on knowledge and data engineering*, 2017, 30(5): 922–935.
- [19] HU Jiafeng, WU Xiaowei, CHENG R, et al. On minimal steiner maximum-connected subgraph queries[J]. *IEEE transactions on knowledge and data engineering*, 2017, 29(11): 2455–2469.
- [20] CHANG Lijun, LIN Xuemin, QIN Lu, et al. Index-based optimal algorithms for computing steiner components with maximum connectivity[C]//Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data. Melbourne: MOD, 2015: 459–474.
- [21] LI Ronghua, SU Jiao, QIN Lu, et al. Persistent community search in temporal networks[C]//2018 IEEE 34th International Conference on Data Engineering. Paris: IEEE, 2018: 797–808.
- [22] RUCHANSKY N, BONCHI F, GARCÍA-SORIANO D,

- et al. To be connected, or not to be connected: that is the minimum inefficiency subgraph problem[C]//Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. New York: CIKM, 2017: 879–888.
- [23] TSALOUCHIDOU I, BONCHI F, BAEZA-YATES R. Adaptive community search in dynamic networks[C]//2020 IEEE International Conference on Big Data. Atlanta: IEEE, 2020: 987–995.
- [24] GUO Tao, CAO Xin, CONG Gao. Efficient algorithms for answering the m-closest keywords query[C]//Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data. New York: MOD, 2015: 405–418.
- [25] WANG Kai, CAO Xin, LIN Xuemin, et al. Efficient computing of radius-bounded k-cores[C]//2018 IEEE 34th International Conference on Data Engineering. Paris: IEEE, 2018: 233–244.
- [26] ZHU Qijun, HU Haibo, XU Cheng, et al. Geo-social group queries with minimum acquaintance constraints[J]. *The VLDB journal*, 2017, 26(5): 709–727.
- [27] CHEN Lu, LIU Chengfei, ZHOU Rui, et al. Maximum co-located community search in large scale social networks[J]. *Proceedings of the VLDB endowment*, 2018, 11(10): 1233–1246.
- [28] LUO Jiehuan, CAO Xin, XIE Xike, et al. Best Co-located community search in attributed networks[C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. New York: CIKM, 2019: 2453–2456.
- [29] LUO Jiehuan, CAO Xin, QU Qiang, et al. Efficient search of the most cohesive Co-located community in attributed networks[C]//International Conference on Database Systems for Advanced Applications. Chiang Mai: Springer, 2019: 398–415.
- [30] FANG Yixiang, CHENG R, LUO Siqiang, et al. Effective community search for large attributed graphs[J]. *Proceedings of the VLDB endowment*, 2016, 9(12): 1233–1244.
- [31] HUANG Xin, LAKSHMANAN L V S. Attribute-driven community search[J]. *Proceedings of the VLDB endowment*, 2017, 10(9): 949–960.
- [32] GHANBARPOUR A, NIKNAFS K, NADERI H. Efficient keyword search over graph-structured data based on minimal covered r-cliques[J]. *Frontiers of information technology & electronic engineering*, 2020, 21(3): 448–464.
- [33] CHOWDHARY A A, LIU Chengfei, CHEN Lu, et al. Finding attribute diversified communities in complex networks[C]//International Conference on Database Systems for Advanced Applications. Jeju: Springer, 2020: 19–35.
- [34] WANG Chunnan, WANG Hongzhi, CHEN Hanxiao, et al. Attributed community search based on effective scoring function and elastic greedy method[J]. *Information sciences*, 2021, 562: 78–93.
- [35] LIU Qing, ZHU Yifan, ZHAO Minjun, et al. VAC: vertex-centric attributed community search[C]//2020 IEEE 36th International Conference on Data Engineering. Dallas: IEEE, 2020.
- [36] LI Ronghua, QIN Lu, YU Jeffreyxu, et al. Influential community search in large networks[J]. *Proceedings of the VLDB endowment*, 2015, 8(5): 509–520.
- [37] LI Ronghua, QIN Lu, YE Fanghua, et al. Skyline community search in multi-valued networks[C]//Proceedings of the 2018 International Conference on Management of Data. Houston: MOD, 2018: 457–472.
- [38] LUO Wensheng, ZHOU Xu, YANG Jianye, et al. Efficient approaches to Top-R influential community search[J]. *IEEE internet of things journal*, 2020, 8(16): 12650–12657.
- [39] BI Fei, CHANG Lijun, LIN Xuemin, et al. An optimal and progressive approach to online search of top-k influential communities[J]. *Proceedings of the VLDB endowment*, 2018, 11(9): 1056–1068.
- [40] CHEN Shu, WEI Ran, POPOVA D, et al. Efficient computation of importancebased communities in web-scale networks using a single machine[C]//ACM International. Indianapolis: ACM, 2016.
- [41] ZHENG Zibin, YE Fanghua, LI Ronghua, et al. Finding weighted K-truss communities in large networks[J]. *Information sciences*, 2017, 417: 344–360.
- [42] XU Jiao, FU Xiaoyi, WU Yiming, et al. Personalized top-n influential community search over large social networks[J]. *World wide web*, 2020, 23(3): 2153–2184.
- [43] XIE Xiaoqin, SONG Mingjie, LIU Chiming, et al. Effective influential community search on attributed graph[J]. *Neurocomputing*, 2021, 444: 111–125.
- [44] SUN Longxu, HUANG Xin, LI Ronghua, et al. Fast algorithms for intimate-core group search in weighted graphs[C]//International Conference on Web Information Systems Engineering. Hong Kong: Springer, 2020: 728–744.
- [45] LIU Boge, YUAN Long, LIN Xuemin, et al. Efficient (α , β)-core computation: an index-based approach[C]//The World Wide Web Conference. San Francisco: ACM, 2019: 1130–1141.

- [46] WANG Kai, LIN Xuemin, QIN Lu, et al. Efficient bitruss decomposition for large-scale bipartite graphs[C]//2020 IEEE 36th International Conference on Data Engineering. Dallas: IEEE, 2020: 661–672.
- [47] ZHANG Yun, PHILLIPS C A, ROGERS G L, et al. On finding bicliques in bipartite graphs: a novel algorithm and its application to the integration of diverse biological data types[J]. *BMC bioinformatics*, 2014, 15(1): 1–18.
- [48] ZHANG Wenqiang, ZHONG Yingli, YANG Yan. Community search in spatial uncertain network[J]. *Journal of physics conference series*, 2021, 1952(4): 042112.
- [49] LUO Wensheng, ZHOU Xu, LI Kenli, et al. Efficient influential community search in large uncertain graphs[J]. *IEEE transactions on knowledge and data engineering*, 2021, 35(4): 3779–3793.
- [50] MIAO Xiaoye, LIU Yue, GAO Yunjun, et al. Reliable community search on uncertain graphs[C]//2022 IEEE 38th International Conference on Data Engineering. Kuala Lumpur: IEEE, 2022.
- [51] 王旭东, 韩晓磊, 邹功鑫, 等. 基于权值的新能源集控电量均衡 CSMA/CA 改进算法 [J]. *微型电脑应用*, 2024, 40(5): 227–230.
WANG Xudong, HAN Xiaolei, ZOU Gongxin, et al. Improved CSMA/CA algorithm of new energy centralized control power balance based on weight[J]. *Microcomputer applications*, 2024, 40(5): 227–230.
- [52] 梁正平, 刘程, 王志强, 等. 基于存档和权值扩展的大规模多目标优化算法 [J]. *计算机学报*, 2022, 45(5): 951–972.
LIANG Zhengping, LIU Cheng, WANG Zhiqiang, et al. Large-scale multi-objective optimization algorithm based on archive and weight extension[J]. *Chinese journal of computers*, 2022, 45(5): 951–972.
- [53] 毛华, 刘祎超. 基于权值最大圈的概念格构造算法 [J]. *智能系统学报*, 2016, 11(4): 519–525.
MAO Hua, LIU Yichao. An algorithm for concept lattice construction based on maximum cycles of weight values[J]. *CAAI transactions on intelligent systems*, 2016, 11(4): 519–525.
- [54] 陈侃松, 李豪科, 阮玉龙, 等. 基于局部邻居节点和链路权值的改进 AODV 路由协议 [J]. *软件学报*, 2021, 32(4): 1186–1200.
CHEN Kansong, LI Haoke, RUAN Yulong, et al. Improved AODV routing protocol based on local neighbor nodes and link weights[J]. *Journal of software*, 2021, 32(4): 1186–1200.
- [55] 王余蓝. 基于图形数据库的 DBLP 数据存储 [J]. *计算机技术与发展*, 2013, 23(8): 18–20, 26.
WANG Yulan. DBLP data storage based on graphics database[J]. *Computer technology and development*, 2013, 23(8): 18–20, 26.

作者简介:



赵丹枫, 副教授, 博士, 主要研究方向为图计算理论、海洋大数据存储、业务流程管理。主持国家自然科学基金青年科学项目, 参与 4 个国家自然科学基金面上项目、上海市地方能力建设项目、海洋科技专项等。发表学术论文 20 余篇。E-mail: dfzhao@shou.edu.cn。



孔万仔, 硕士研究生, 主要研究方向为图计算理论、数据库。E-mail: kwzshou@163.com。



刘国华, 教授, 主要研究方向为数据库理论、隐私保护和 BPMS。以主要参加人的身份参与过 1 项国家自然科学基金项目和多项省自然科学基金项目的研究工作, 发表学术论文 100 余篇, 出版专著 3 部。E-mail: ghliu@dhu.edu.cn。