

簇图编辑问题的研究进展

王建新 万茂武 陈建二

(中南大学信息科学与工程学院 长沙 410083)

摘 要 簇图编辑问题是一个重要的 NP-难问题。作为相关性聚类问题的一个特例,它在计算生物等领域有着重要的应用。参数计算理论出现后,参数化的簇图编辑问题逐渐引起了很多人的注意。介绍了求解簇图编辑问题的近似算法、参数算法和它的一些变形,着重分析了参数化簇图编辑问题核心化和 FPT 算法的最新结果。最后提出了关于该问题的一些研究方向。

关键词 簇图编辑问题,聚类问题, NP-难,近似算法,参数算法

中图法分类号 TP301.6 **文献标识码** A

Algorithms for Cluster Editing: A Survey

WANG Jian-xin WAN Mao-wu CHEN Jian-er

(School of Information Science and Engineering, Central South University, Changsha 410083, China)

Abstract The Cluster Editing is known to be a very important NP-hard problem. As a special case of Correlation Clustering, it plays a significant role on many fields such as computation biology. After the theory of parameterized complexity was brought up, its parameterized version has drawn much attention. We introduced some approximation and parameterized algorithms for this problem and some variants of it, emphasized on the latest results about its kernelization and FPT algorithms. At the end, we presented some further directions for future research.

Keywords Cluster editing, Clustering, NP-hard, Approximation algorithms, Parameterized algorithms

1 引言

簇图编辑问题是在给定的图中通过对图进行删边和添边操作使其变为簇图(簇图的定义见本文第 2 节),要求删除和添加的边尽可能少。

簇图编辑问题是一种特殊的聚类问题,在计算生物^[1,2]、机器学习^[3,4]等领域都有重要应用。以机器学习中的文档分类问题为例,假设现有 n 个文档需要按主题进行分类,但是事先并没有给定有哪些主题。一种常见的做法是首先通过在以往的数据集上进行训练,学习如何判断任意两个文档在主题上是否相似。然后利用学到的分类知识处理这 n 个文档,判断每两个文档是否相似。最后对这些文档进行聚类,使得聚类后的结果与前面处理的结果尽可能一致。对上述问题利用图形进行建模即可转化为簇图编辑问题。计算生物中很多聚类问题也都可以转化为簇图编辑问题。

簇图编辑问题已经证明是 NP-完全问题。现有的研究主要基于近似算法和参数化算法,并取得了很好的结果。特别是近年来,很多文献都在研究该问题的参数化算法。在此对其进行总结,希望人们对该问题的研究有更全面的了解,也有利于我们在该问题上进行更深入的研究。

本文第 2 节介绍簇图编辑问题的多项式时间近似算法;

第 3 节介绍参数化算法,重点分析算法中使用的分枝搜索和核心化这两种技术;第 4 节介绍一些变形问题的研究成果;最后总结簇图编辑问题的研究现状并展望未来的研究工作。

2 近似算法

簇图编辑问题的具体定义如下:

定义 1(簇图编辑问题, Cluster Editing) 给定一个图 $G=(V, E)$, 设图 G 的补图为 $G'=(V, E')$ 。试求两个边集合 $E_d \subseteq E$ 和 $E_a \subseteq E'$, 使得图 $G''=(V, E-E_d+E_a)$ 的每一个连通分量都是一个完全连通图并且 $|E_d|+|E_a|$ 最小(若一个图的每一个连通分量都是一个完全连通图, 则称该图是一个簇图)。

簇图编辑问题被证明是 NP-难的^[4-7]。为此,人们采用近似算法对其进行求解。表 1 列出了簇图编辑问题近似算法的主要研究进展。

表 1 簇图编辑问题的近似算法比较

文献	相对近似率
[3,4]	$9(1/\delta^2+1)$, $\delta=1/44$
[8]	APX-难
[8]	4
[9,10]	3(随机近似)
[11,12,13]	2.5

到稿日期:2010-01-26 返修日期:2010-04-12 本文受国家自然科学基金(60773111)和国家教育部创新团队资助计划(IRT0661)资助。

王建新 男,博士,教授,博士生导师,CCF 高级会员,主要研究方向为计算机算法、网络优化理论、生物信息学, E-mail: jxwang@mail.csu.edu.cn; 万茂武 男,硕士生,主要研究方向为参数计算; 陈建二 男,博士,教授,博士生导师,主要研究方向为生物信息学、计算理论、计算复杂性及优化。

Bansal 等在文献[3,4]中主要研究了相关性聚类问题(Correlation Clustering)的近似算法。簇图编辑问题是相关性聚类问题在输入是完全连通图时的特例。定义1中的簇图编辑问题优化的目标是 $(|E_d| + |E_a|)$ 最小,与此等价的目标是 $(|E| + |E'| - (|E_d| + |E_a|))$ 最大。文献[3,4]设计的近似算法针对这两个目标分别提出了不同的近似算法。对于前者,提出了一个相对近似率为 $9(1/\delta^2 + 1)$ 的近似算法(其中 $\delta=1/44$)。对于后者,证明了它是 PTAS 的(即对于任意给定的常数近似率,总存在一个多项式时间近似算法满足给定的近似率)。

Charikar 等^[8]以“使 $(|E_d| + |E_a|)$ 最小”为目标给出了一个相对近似率为4的近似算法,从而大大改进了 Bansal 的结果。该近似算法基于线性规划松弛法(linear programming relaxation),使用区域增长法(region growing)对非整数解进行取整。作为补充,Charikar 等^[8]证明了簇图编辑问题是 APX-难的,即总存在一个常数 $\epsilon > 0$,簇图编辑问题不存在一个相对近似率为 $(1+\epsilon)$ 的近似算法,除非 $P=NP$ 。可见,对于簇图编辑问题,以“使 $(|E_d| + |E_a|)$ 最小”为目标进行近似和以“使 $(|E| + |E'| - (|E_d| + |E_a|))$ 最大”为目标进行近似并不等价,前者是 APX-难,而后者则是 PTAS。

Ailon 等在文献[9,10]中主要研究了排名(ranking)和聚类(clustering)问题。文章从锦标赛图上的反馈边集问题出发,提出了一个类似于随机快速排序的随机算法,然后将这个随机算法扩展到排名聚合问题(Ranking Aggregation)、相关性聚类 and 一致性聚类问题(Consensus Clustering)上。在分析算法的相对近似率时,引入了“坏三元组”概念(bad triplet),并运用概率方法证明算法的期望相对近似率为3。

基于 Ailon 的研究成果,Zuylen 等^[11-13]采用摊销(amortize)的思想对 Ailon 的随机算法进行确定化,确定化后的算法的相对近似率为2.5。这是目前求解簇图编辑问题相对近似率最好的一个近似算法。

3 参数算法

在某些实际应用中,人们常常需要获得精确解,这时近似算法无法满足人们的要求。近年来出现了一种求解 NP-难问题的新思路——参数算法。该理论根据工程应用的实际对原问题进行参数化,充分利用实际工程应用中“小参数”的特点,设计快速有效的精确算法。参数算法将计算理论和计算实践有效地结合在一起。

对于簇图编辑问题,通常选择在原图中删除和添加的边的总数作为参数 k ,从而得到如下参数问题。

定义2(参数化的簇图编辑问题, Cluster k -Editing) 给定一个图 $G=(V, E)$ 和非负整数 k ,设图 G 的补图为 $G'=(V, E')$ 。试求两个边集合 $E_d \subseteq E$ 和 $E_a \subseteq E'$,使得图 $G''=(V, E-E_d+E_a)$ 是簇图并且 $|E_d| + |E_a| \leq k$ 。

Cai 在文献[14]中研究了更一般的图修改问题,证明了如果修改后的图类具有继承性,并且该继承性可通过有限个极小的非法导出子图(minimal forbidden induced subgraph)来刻画,那么以修改操作的次数作为参数,对应的参数化图修改问题是固定参数可解的(Fixed Parameter Tractable, FPT)。利用 Cai 的这一结论可知 Cluster k -Editing 问题是 FPT 的,并且按照 Cai 在文献[14]中的方法,对图中含有的 P_3 (一条

长度为2的路径)导出子图进行分枝,可以得到一个时间复杂度为 $O(3^k n^4)$ 的固定参数算法。

对于固定参数可解问题,通常的研究方向包括设计更有效的固定参数算法和获得更小的核。在参数算法设计技术中,核化是一种重要的技术,通过利用预处理规则可以在多项式时间内降低问题输入实例的规模。Chen 等在文献[15]中指出,固定参数可解和可核化是等价的,因此 Cluster k -Editing 问题也是可核化的。

表2列出了 Cluster k -Editing 问题的参数化算法和核化结果。目前,最好的参数化算法的时间复杂度为 $O(1.82^k + n^3)$,最好的核化结果是大小为 $4k$ 的核(以下若无特别说明,核的大小均是按照图中的顶点数目计算)。

表2 Cluster k -Editing 问题的参数化算法比较

文献	FPT 算法	核大小	核化时间
[14]	$O(3^k n^4)$	—	—
[16,17]	$O(2.27^k + n^3)$	$O(k^2)$	$O(n^3)$
[18]	$O(1.92^k + n^3)$	—	—
[19,20]	$O(1.82^k + n^3)$	—	—
[21]	—	$O(k^2)$	$O(n+m)$
[22]	—	24k	$O(n^3)$
[23]	—	6k	$O(n^3)$
[24,25]	—	4k	$O(nm^2)$

3.1 分枝搜索

在利用分枝搜索技术设计参数算法时,算法的执行过程可用一棵分枝搜索树来表示;树根是输入的实例或者经过核化后的实例,对于每一个树结点,在多项式时间内找出图中的一个子结构,使得该子结构中至少有一个元素在问题的最优解中。然后对该子结构进行分枝:每个分枝从子结构中选出一个不同的元素放入最优解中,这时若以最优解的大小作为参数 k ,则 k 在每个分枝上至少减1。如此分枝直到参数 k 小于0或者图中已不存在这样的子结构。

整个算法的运行时间取决于分枝搜索树的规模(通常用树中结点总数或者叶子结点总数来进行度量)和在每个树结点上所花费的时间(这个时间通常是多项式时间)。若设在某个树结点上的参数值为 t ,进行分枝后得到的每一个孩子结点的参数值分别是 $t-d_1, t-d_2, \dots, t-d_i$,则称 $(d_1, d_2, \dots, d_i)$ 是此次分枝的分枝向量。设分枝搜索树的树根上的参数值为 k 并且树中每一个内部结点都按照分枝向量 $(d_1, d_2, \dots, d_i)$ 进行分枝,可以证明整个分枝搜索树的大小是 $O(n^{\alpha(1)} \cdot |\alpha|^k)$,其中 α 是特征方程 $\alpha^k = \alpha^{k-d_1} + \alpha^{k-d_2} + \dots + \alpha^{k-d_i}$ 的根^[26](称 α 是分枝向量 $(d_1, d_2, \dots, d_i)$ 对应的分枝数)。由此可见,要大幅度降低算法的运行时间,需要尽可能减少分枝搜索树的规模,而这通常需要更巧妙的分枝策略。

对于簇图编辑问题,由于一个图是簇图的充要条件是图中不含有 P_3 这个导出子图,因此可以对图中的任意一个 $P_3 = \{u, v, w\}$ (其中 $uv \in E, uw \in E, vw \notin E$) 进行分枝:(B1)删除边 uv ;(B2)删除边 uw ;(B3)添加边 vw 。这种分枝规则对应的分枝向量是 $(1, 1, 1)$,相应的分枝数为3。因此可得到一个时间复杂度为 $O^*(3^k)$ 的固定参数算法^[14]。

J. Gramm 等发现若 $P_3 = \{u, v, w\}$ (其中 $uv \in E, uw \in E, vw \notin E$) 满足:点 v 和 w 除了点 u 这个共同邻居之外再无其它共同邻居,则无需产生“(B3)添加边 vw ”这个分枝。对这种特殊情况的 P_3 ,对应的分枝向量是 $(1, 1)$,相应的分枝数是2。

此外,还对两个特殊的包含 P3 的大小为 4 的导出子图进行分枝,对应的分枝向量分别是(1,2,3,2,3)和(1,2,3,3,2),相应的分枝数都是 2.27。取最坏情况作为算法的时间复杂度,从而得到了一个时间复杂度为 $O^*(2.27^k)$ 的固定参数算法^[16,17]。

在之前工作的基础上,J. Gramm 等提出了分枝搜索自动化的概念。首先找出一系列规模较小的包含 P3 这个导出子图的子结构,然后利用元分枝搜索树为每一个子结构找出最优的分枝向量,最后在分析分枝搜索树的规模时只需按照具有最大分枝数的分枝向量进行计算^[18]。J. Gramm 等运用分枝搜索自动化找出了 137 个大小为 6 的待分枝子结构,最大的分枝向量有 37 个分枝,最大的分枝数是 1.92,从而得到了一个时间复杂度为 $O^*(1.92^k)$ 的固定参数算法^[18],极大地改进了之前的算法。通过为更大的待分枝子结构设计分枝策略,预计这个时间复杂度还可以进一步改进。值得一提的是,这种分枝搜索自动化的方法可以推广到更一般的图修改问题上^[18]。

最近,S. Böcker 等^[19,20]研究了参数化带权簇图编辑问题。每删除或添加一条边都需要付出一定的代价,问题的参数 k 是总的编辑代价,通过运用一系列更加精细、更加巧妙的分枝策略,将分枝搜索树的规模一降再降,从 $O(3^k)$ 到 $O(2.42^k)$,再到 $O(2^k)$,最后到 $O(1.82^k)$ 。该分枝搜索算法也可直接应用到不带权的情况(每次编辑的代价是 1),从而得到簇图编辑问题的一个时间复杂度为 $O^*(1.82^k)$ 的固定参数算法^[19,20]。这是目前求解簇图编辑问题时间复杂度最小的参数算法。

需要说明的是,理论上时间复杂度最小的算法在变为具体程序后并不意味着运行速度最快,因为越是精细巧妙的分枝策略,也就意味着更多、更繁琐的分枝规则,实现这些繁琐的规则将要花费很大的开销。

3.2 核心化

核心化是参数计算理论中最重要的概念之一,通过一些预处理规则可以大幅度降低问题输入实例的规模^[27]。为参数化问题找出更小的核,一直是研究者们追求的目标。簇图编辑问题的核也经过研究者的努力从平方核降到了线性核。

J. Gramm 等^[16,17]提出了 3 条很容易验证的规则:(1)若一个点对 u 和 v 相同的邻居数超过 k ,则 u 和 v 之间必须有边;若它们相异的邻居数超过 k ,则 u 和 v 之间不能有边;若相同和相异的邻居数都超过 k ,则输入实例无解。(2)若 u 和 v 之间、 u 和 w 之间都必须有边,则 v 和 w 之间也必须要有边;若 u 和 v 之间必须有边, u 和 w 之间不能有边,则 v 和 w 之间也不能有边。(3)直接删除图中完全连通的连通分量。经过复杂的分析过程,最后将原实例降到平方核,所用时间为 $O(n^3)$ 。

F. Protti 等在文献[21]中运用“模块”(Module)概念,将原图进行模块分解,得到了与原图等价的 p -商图 G_p 。在 G_p 中有两类顶点: p -顶点和 u -顶点。其中 p -顶点对应原图中的一个顶点数大于 1 的独立集模块,而 u -顶点则直接对应原图中的一个点。假设原图存在一个大小不超过 k 的解(或极小解),若解中存在这样一条边 (a, v_a) ,其中 a 属于 G_p 中某个 p -顶点所对应的模块 M_a ,则对于 M_a 中其他任意一点 b ,边 (b, v_a) 也必在解中,因为点 a 和点 b 对于模块 M_a 外的点而言是对称的。换言之,如果 G_p 中某个 p -顶点所对应的模块 M 大小超过 k ,则 M 中的任何一点都不可能是解中某条边的端

点。因此可以按照如下方法构造与原图等价的一个核 $G_k = (V_k, E_k)$:将 G_p 中的 u -顶点直接放入 V_k 中,对于 G_p 中的 p -顶点,如果它所对应的模块大于 k ,则只需要用其中的任意 $(k+1)$ 个顶点来代替,将它们放入 V_k 中;如果不超过 k ,则直接放入 V_k 中,而 G_k 只是由 V_k 导出的子图。通过限制图 G_p 中的 p -顶点数目和 u -顶点数目,可以证明按照上述方法得到的核含有 $2k(k+1)+2k=2k^2+4k$ 个顶点,并且上述构造过程只需要花费线性时间。

M. Fellows 等将顶点覆盖问题(Vertex Cover)线性核心化过程中运用的皇冠分解技术(Crown Decomposition)进行扩展,定义了簇图编辑问题下的新皇冠,提出了一条簇皇冠简化规则(Cluster Crown Reduction Rule),从而大幅度地减少了原图的规模。在一份手稿中,证明了该规则的正确性,并证明原图运用该规则后最多只包含 24k 个点^[22]。后来,在不增加新规则的基础上这个核的上界又被进一步降到 $6k$ ^[23]。

最近,J. Guo 通过运用关键团(Critical Clique)的概念,将原图转化为关键团图。所谓关键团是指这样的极大团,团中任意两点只有共同的邻居而无各自私有的邻居。在关键团图的基础上提出了 4 条核心化规则,其中第二条规则其实就是 M. Fellows 等提出的簇皇冠简化规则。J. Guo 证明了运用第一、二条规则可以得到大小为 $6k$ 的核,运用第一、三、四条规则可以得到大小为 $4k$ 的核,两个核心化过程所用的时间分别是 $O(n^3)$ 和 $O(nm^2)$ ^[24,25]。 $4k$ 大小的核是目前求解簇图编辑问题核心化过程的最好结果。

关于簇图编辑问题,F. Dehne 和 S. Böcker 等分别进行了相关算法的工程实现,并通过实验验证了一些核心化规则和分枝策略的高效性。具体实现过程参见文献[28,29]。

4 问题的变形

簇图编辑问题有很多变形,大致可以分为 3 类:(1)编辑操作的变形;(2)簇图条件的变形;(3)枚举。下面分别介绍这 3 类变形所涉及的一些问题的研究成果。以下若未作特别说明,各参数化问题均以问题解的大小作为参数 k 。

4.1 编辑操作的变形

在簇图编辑问题中,编辑操作包括删除边和添加边。其他允许的编辑操作还有:只允许删除边,只允许添加边和只允许删除点(只添加点是无法从非簇图变为簇图的)。对应这 3 种操作,分别有 3 个问题:簇图删边问题(Cluster Deletion)、簇图添边问题(Cluster Completion)和簇图删点问题(Cluster Vertex Deletion)。这 3 个问题中,只有簇图添边问题是多项式时间可解的^[7],其余两个均是 NP-难的^[7,30]和 APX-难的^[7,31]。从参数复杂性角度看,簇图删边问题和簇图删点问题均是 FPT 的^[14]。对于簇图删边问题,目前最好的参数算法是由 J. Gramm 等提出的时间复杂度为 $O(1.53^k + n^3)$ 的算法^[18]。对于簇图删点问题,目前最好的参数算法是由 F. Hüffner 等提出的时间复杂度为 $O(2^k \cdot k^6 \log k + nm)$ 的算法,最小的核是基于 3-Hitting Set 问题的大小为 $O(k^2)$ 的核^[32]。

4.2 簇图条件的变形

簇图编辑问题要求最终得到的图中每一个簇均是一个完全连通图。簇图这个定义的限制条件过于严格,对其适当放宽可以得到很多其他类型的簇图,比如双簇图、 s -丛图等。对应这两个簇图概念,分别有双簇图编辑问题(Bicluster Editing)和 s -丛图编辑问题(s -Plex Editing)。

双簇图(Biclust)的定义:如果一个二部图 $G=(L,R;E)$ 满足 L 中的每一点和 R 中的每一点都有边相连,则称图 G 是双簇图。双簇图编辑问题是指通过删除和添加最少的边使得图中每一个连通分量都是一个双簇图。它是 NP-难的^[33],也是 FPT 的^[14]。目前最好的近似算法是期望近似率为 4 的随机近似算法^[34],最小的核大小为 $6k$ ^[34],最好的参数算法时间复杂度为 $O(3.24^k+m)$ ^[34]。

s -丛图(s -Plex)的定义:如果一个连通图 $G=(V,E)$ 满足 V 中每一个顶点的度数均不小于 $|V|-s$,那么称图 G 是一个 s -丛图,当 $s=1$ 时,1-丛图就是完全连通图。 s -丛图编辑问题是指通过删除和添加最少的边使得图中每一个连通分量都是一个 s -丛图,当 $s=1$ 时,此问题退化为簇图编辑问题。对于常数 $s \geq 2$, s -丛图编辑问题是 NP-难的^[35],也是 FPT 的^[35]。目前该问题有一个大小为 $(4s^2-2)k+4(s-1)^2$ 的核^[35],通过对非法子结构进行分枝可以得到一个时间复杂度为 $O((2s+\lfloor s^{1/2} \rfloor)^k+n^4)$ 的参数算法^[35]。

簇图编辑问题中要求得到的图中各簇之间没有重叠,放松这个条件,允许各簇在某些点或边上重叠,这样可以定义一系列新的问题。

s -重点(边)簇图编辑问题(s -Vertex(Edge) Overlap Editing):给定图 $G=(V,E)$,试删除和添加最少的边使得图中每一个顶点(边)至多在 s 个极大团中。对于常数 $s \geq 1$,此问题是 NP-难的^[36],对应的参数问题是 $W^{[1]}$ -难的^[36]。如果以 s 和 k 的组合作为参数,则该问题是 FPT 的^[36]。特别地,对于 1-重边簇图编辑问题目前有一个大小为 $O(k^4)$ 的核^[36];对于 2-重点簇图编辑问题目前有一个大小为 $O(k^3)$ 的核^[36]。

另外,如果限制簇图中簇的个数,则会得到 d -簇图编辑(d -Cluster Editing)问题:给定图 $G=(V,E)$,试删除和添加最少的边使得图中有且仅有 d 个完全连通图。当 $d=1$ 时,此问题多项式时间可解;当 $d \geq 2$ 且固定,此问题是 NP-难的^[7]。对于该问题的参数化形式,J. Guo 在 $O(m+n)$ 时间内给出了一个大小为 $(d+2)k+d$ 的核^[24,25],从而说明该问题也是 FPT 的。

4.3 枚举

上面所研究的簇图编辑问题都只需要求出问题的一个解(如果有解的话)。在实际应用中,枚举出所有的解也是很有意义的。P. Damaschke 在文献^[37]中研究了参数化簇图编辑问题的枚举,即给定图 $G=(V,E)$ 和非负整数 k ,通过删除和添加不超过 k 条边使得图 G 变为一个簇图,试枚举出所有的极小解(即任何两个解中一个不会包含另一个)。对于该问题,给出了一个时间复杂度为 $O(2.27^k+k^2n+m)$ 的固定参数算法,从而证明它也是 FPT 的。

除了上面提到的 3 种变形之外,它们的组合也能产生许多新的问题,例如 s -丛图删点问题(s -Plex Vertex Deletion)、双簇图删边问题的枚举(Enumeration of Biclust Deletion)等。总之,对簇图条件进行适当变形,可以得到很多有意义的问题,它们在计算生物学中有很多应用。

结束语 簇图编辑问题是一个很有意义的 NP-难问题,在很多领域中都有应用,尤其是在计算生物中。人们对它做了大量的研究,总结起来,大致可以分为以下 3 个方面:

(1)研究多项式时间近似算法。对于 NP-难问题,很难通过精确算法有效地求出大规模实例的解,这时以牺牲精度为代价来换取时间,也是一种很可行的办法。

(2)研究参数算法。参数复杂性理论为求解 NP-难问题提供了新的思路。通过核心化规则可以大幅度降低实例的规模,利用固定参数算法并结合“小参数”的特点,可以精确求解实际工程中的很多实例。

(3)研究问题的变形。求解变形后的新问题,证明其难度和参数复杂性,给出近似算法或者参数算法。

由于簇图编辑问题来源于计算生物这个新兴的学科,对于它的研究才刚刚起步,很多地方还不完善。进一步的研究可以考虑以下几个方面:

(1)关于簇图编辑问题及其变形有很多核心化规则和算法,目前它们大多只是停留在理论结果上。鉴于这些问题在实际工程应用中的重要性,对这些算法进行软件实现,以验证其实际效果,将变得很有意义,而目前这方面的工作还很少。

(2)进一步降低某些问题核的大小,尤其是簇图编辑问题的变形。目前人们对这些变形问题的研究才起步,很多地方有待进一步挖掘。例如簇图删点问题目前只有平方核,能否给出它的一个线性核呢?

(3)进一步降低某些算法或者核心化过程的时间复杂度。例如尽管簇图编辑问题目前已有大小为 $4k$ 的核,但是核心化时间复杂度为 $O(nm^2)$,能否将它降到线性时间呢?

(4)提出簇图编辑问题新的变形,并尝试用近似算法、参数算法等对其进行求解。例如对簇重叠的概念进行扩展,可以得到很多新的问题。

参考文献

- [1] Ben-Dor A, Shamir R, Yakhini Z. Clustering Gene Expression Patterns[J]. Journal of Computational Biology, 1999, 6(3/4): 281-297
- [2] Witkop T, Baumbach J, Lobo F, et al. Large Scale Clustering of Protein Sequences with FORCE—A Layout Based Heuristic for Weighted Cluster Editing[J]. BMC Bioinformatics, 2007, 8(1): 396
- [3] Bansal N, Blum A, Chawla S. Correlation Clustering[C]// Proc. 43rd FOCS. 2002; 238-247
- [4] Bansal N, Blum A, Chawla S. Correlation Clustering [J]. Machine Learning, 2004, 56(1-3): 89-113
- [5] Krivánek M, Morávek J. NP-Hard Problems in Hierarchical Tree Clustering[J]. Acta Informatica, 1986, 23(3): 311-323
- [6] Chen Zhi-zhong, Jiang Tao, Lin Guohui. Computing Phylogenetic Roots with Bounded Degrees and Errors[J]. SIAM Journal on Computing, 2003, 32(4): 864-879
- [7] Shamir R, Sharan R, Tsur D. Cluster Graph Modification Problems[J]. Discrete Applied Mathematics, 2004, 144: 173-182
- [8] Charikar M, Guruswami V, Wirth A. Clustering with Qualitative Information[J]. Journal of Computer and System Sciences, 2005, 71: 360-383
- [9] Ailon N, Charikar M, Newman A. Aggregating Inconsistent Information: Ranking and Clustering [C] // Proc. 37th ACM STOC. ACM Press, 2005; 684-693
- [10] Ailon N, Charikar M, Newman A. Aggregating Inconsistent Information: Ranking and Clustering[J]. Journal of ACM, 2008, 55(5): 23
- [11] van Zuylen A, Hegde R, Jain K, et al. Deterministic Pivoting Algorithms for Constrained Ranking and Clustering Problems[C]// Proc. 8th Annual ACM-SIAM Symposium on Discrete Algorithms. 2007; 405-414

参考文献

- [1] 孙利民, 李建中, 陈渝, 等. 无线传感器网络[M]. 北京: 清华大学出版社, 2005
 - [2] Akyildiz I F, Su W, Sankarasubramaniam Y, et al. Wireless sensor networks: A survey[J]. Computer Networks, 2002, 38(4): 393-422
 - [3] 王亮, 钟先信, 石军锋. 无线传感器网络能耗平衡策略的研究[J]. 传感器世界, 2007(3): 32-36
 - [4] 吴小兵, 陈贵海. 无线传感器网络中节点非均匀分布的能量空洞问题[J]. 计算机学报, 2008, 31(2): 253-261
 - [5] Luo J, Hubaux J P. Joint Mobility and Routing for Lifetime Elongation in Wireless Sensor Networks[C]//Proc. INFOCOM, Mar. 2005
 - [6] Lian J, Chen L, Naik K, et al. Modeling and enhancing the data capacity of wireless sensor networks[C]//Phoha S, La Porta T F, Griffin C, eds. IEEE Monograph on Sensor Network Operations. IEEE Press, 2004: 91-138
 - [7] Song L, Hatzinakos D. Architecture of wireless sensor networks with mobile sinks; sparsely deployed sensors [J]. IEEE Transactions on Vehicular Technology, 2007, 56(4): 1826-1836
 - [8] Li J, Mohapatra P. An analytical model for the energy hole problem in many-to-one sensor networks[C]//Proceedings of IEEE Vehicular Technology Conference, Dallas, TX, 2005: 2721-2725
 - [9] 赵文辉, 刘大昕, 姜宇, 等. 无线传感器网络节点分布策略研究[J]. 计算机工程与应用, 2008, 44(35): 115-118
 - [10] 羊四清, 袁辉勇, 易叶青. 无线传感器网络中非均匀的节点布置[J]. 计算机工程与应用, 2008, 44(24): 118-120
 - [11] Wang W, Srinivasan V, Chua K. Using mobile relays to prolong the lifetime of wireless sensor networks[C]//Proceedings of ACM MobiCom, Cologne, Germany, 2005: 270-283
 - [12] 石高涛, 廖明宏. 传感器网络中具有负载均衡的移动协议数据收集模式[J]. 软件学报, 2007, 18(9): 2235-2244
 - [13] Gandham S R, Dawande M, Prakash R, et al. Energy Efficient Schemes For Wireless Sensor Networks with Multiple Mobile Base Station[C]//The IEEE Global Telecommunication Conference(GLOBECOM). 2003
 - [14] 程龙, 陈灿峰, 马建. 无线传感器网络中多移动 sink 的选择策略[J]. 通信学报, 2008, 29(11): 12-19
 - [15] Heinzelman W, Chandrakasan A, Balakrishnan H. An application-specific protocol architecture for wireless microsensor networks [J]. IEEE Transactions on Wireless Communications, 2002, 1(4): 660-670
 - [16] Ye Mao, Li Chengfa, Chen Guihai, et al. EECS: An Energy Efficient Clustering Scheme in Wireless Sensor Networks[C]//IEEE Int. Performance Computing and Communication Conference. 2005: 535-540
-
- (上接第 11 页)
- [12] van Zuylen A, Williamson D P. Deterministic Algorithms for Rank Aggregation and Other Ranking and Clustering Problems [J]. LNCS, 2008, 4927: 260-273
 - [13] van Zuylen A, Williamson D P. Deterministic Pivoting Algorithms for Constrained Ranking and Clustering Problems[J]. Mathematics of Operations Research, 2009, 34(3): 594-620
 - [14] Cai Leizhen. Fixed-parameter Tractability of Graph Modification Problems for Hereditary Properties[J]. Information Processing Letters, 1996, 58: 171-176
 - [15] Chen Jianer. Parameterized Computation and Complexity: A New Approach Dealing with NP-hardness[J]. Journal of Computer Science and Technology, 2005, 20: 18-37
 - [16] Gramm J, Guo J, Hüffner F, et al. Graph-modeled Data Clustering: Fixed-parameter Algorithms for Clique Generation[C]//Proc. 5th CIAC(2003). LNCS 2653, 2003: 108-119
 - [17] Gramm J, Guo J, Hüffner F, et al. Graph-modeled Data Clustering: Fixed-parameter Algorithms for Clique Generation [J]. Theory of Computing Systems, 2005, 38: 373-392
 - [18] Gramm J, Guo J, Hüffner F, et al. Automated Generation of Search Tree Algorithm for Hard Graph Modification Problems [J]. Algorithmica, 2004, 39: 321-347
 - [19] Böcker S, Briesemeister S, Bui Q B A. Going Weighted: Parameterized Algorithms for Cluster Editing[C]//Proc. 2nd COCOA (2008). LNCS 5165, 2008: 1-12
 - [20] Böcker S, Briesemeister S, Bui Q B A. Going Weighted: Parameterized Algorithms for Cluster Editing[J]. Theoretical Computer Science, 2009, 410: 5467-5480
 - [21] Protti F, da Silva M D, Szwarcfiter J L. Applying Modular Decomposition to Parameterized Bicluster Editing[C]//Proc. 2nd IWPEC(2006). LNCS 4169, 2006: 1-12
 - [22] Fellows M R, Langston M A, Rosamond F, et al. Polynomial-Time Linear Kernelization for Cluster Editing[Z]. Manuscript, 2006
 - [23] Fellows M R, Langston M A, Rosamond F, et al. Efficient Parameterized Preprocessing for Cluster Editing[C]//Proc. 16th FCT(2007). LNCS 4639, 2007: 312-321
 - [24] Guo J. A More Effective Linear Kernelization for Cluster Editing [C]//1st ESCAPE(2007). LNCS 4614, 2007: 36-47
 - [25] Guo J. A More Effective Linear Kernelization for Cluster Editing [J]. Theoretical Computer Science, 2009, 410: 718-726
 - [26] Niedermeier R. Invited to Fixed-Parameter Algorithms[M]. Oxford University Press, 2006
 - [27] Hüffner F, Niedermeier R, Wernicke S. Techniques for Practical Fixed-parameter Algorithms[J]. The Computer Journal, 2008, 51(1): 7-25
 - [28] Dehne F, Langston M A, Luo X, et al. The Cluster Editing Problem: Implementations and Experiments[C]//Proc. 2nd IWPEC (2006). LNCS 4169, 2006: 13-24
 - [29] Böcker S, Briesemeister S, Klau G W. Exact Algorithms for Cluster Editing: Evaluation and Experiments [C]//Proc. 7th WEA(2008). LNCS 5038, 2008: 289-302
 - [30] Lewis J M, Yannakakis M. The Node-deletion Problem for Hereditary Properties is NP-complete[J]. Journal of Computer and System Sciences, 1980, 20(2): 219-230
 - [31] Lund C, Yannakakis M. The Approximation of Maximum Subgraph Problems[C]//Proc. 20th ICALP. LNCS 700, 1993: 40-51
 - [32] Abu-Khzam F N. Kernelization Algorithms for d-Hitting Set Problems[C]//Proc. 10th WADS. LNCS 4619, 2007: 434-445
 - [33] Amit N. The Bicluster Graph Editing Problem [D]. School of Mathematical Sciences, Tel Aviv University, 2004
 - [34] Guo J, Hüffner F, Komusiewicz C, et al. Improved Algorithms for Bicluster Editing[C]//Proc. 5th TAMC. LNCS 4978, 2008: 445-456
 - [35] Guo J, Komusiewicz C, Niedermeier R, et al. A More Relaxed Model for Graph-based Data Clustering: s-Plex Editing[C]//Proc. 5th AAIM. LNCS 5564, 2009: 226-239
 - [36] Fellows M R, Guo J, Komusiewicz C, et al. Graph-based Data Clustering with Overlaps[C]//Proc. 15th COCOON(2009). LNCS 5609, 2009: 516-526
 - [37] Damaschke P. On the Fixed-parameter Enumerability of Cluster Editing[C]//Proc. 31st WG(2005). LNCS 3787, 2005: 283-294