

最大圈分解问题的研究进展

罗卫东 王建新 冯启龙

(中南大学信息科学与工程学院 长沙 410083)

摘要 最大圈分解问题最早由 Erdős 和 Pósa 提出,随后研究人员在图论领域和理论计算机科学领域中对其进行了广泛的探索。最近研究发现,该问题在计算生物学上特别是在构建进化树与分析基因组的研究方面有重要的应用。主要介绍了该问题的研究现状。首先讨论了该问题在图论方面的研究进展;随后对该问题的近似算法、参数算法、参数复杂性与不可近似性进行了分析和讨论;最后给出了该问题的进一步研究方向。

关键词 最大圈分解,图论,近似算法,参数算法,不可近似性,参数复杂性

中图法分类号 TP301.6 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.01.001

Survey of Cycle Packing Problem

LUO Wei-dong WANG Jian-xin FENG Qi-long

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

Abstract Cycle packing problem was first presented by Erdős and Pósa. Since then researchers have been studying the problem in graph theory and theoretical computer science. Recently, researchers find that the problem has important applications in computational biology, especially for reconstruction of evolutionary trees and genomic analysis. In this paper, an introduction to the research status of this problem was given. First and foremost, some results of cycle packing problem in graph theory were discussed. Then, we analyzed and discussed approximation algorithms, parameterized algorithms, parameterized complexity and inapproximability of the problem. At last, some further research directions on this problem were given.

Keywords Cycle packing, Graph theory, Approximation algorithms, Parameterized algorithms, Inapproximability, Parameterized complexity

1 引言

在计算生物学领域中,最大圈分解问题有非常多的应用^[47-51,54],例如,求两个基因组之间的编辑距离问题等价于在这两个基因组生成的邻接点图中求最大圈分解的问题。Erdős 等人最早提出了最大圈分解问题,并且给出了任意一个图中的最大圈分解与该图的顶点反馈集之间的关系,随后出现了大量该问题图论方面的研究。随着该问题在计算生物学方面的应用,问题算法方面的研究也开始受到人们的重视,得到了该问题的各类算法方面与计算复杂性方面的结论。

首先介绍各种类型的最大圈分解问题。最大圈分解问题(Cycle Packing, CP)是指给定一个图 G (图 G 为有向图或无向图),图 G 中有一些满足某种特性的圈,这些圈组成集合 F ,求集合 F 中元素个数的最大值。当集合 F 中的圈满足不同的性质时,对应不同类型的最大圈分解问题。例如,若要求 F 中的圈两两之间没有公共顶点,则该问题为点不相交最大圈分解问题(Vertex-Disjoint Cycle Packing, VDCP);若要求 F 中的圈两两之间没有公共边,则该问题为边不相交最大圈分解问题(Edge-Disjoint Cycle Packing, EDCP)。设集合 S 为 V 的子集,若 G 中某个圈的点构成的集合为 C ,且 C 与 S 的交

集不为空,则称该圈为 S 圈。若要求 F 中的所有圈都为 S 圈,该问题称为最大 S 圈分解问题。最大 S 圈分解问题也可以根据点不相交与边不相交两类,分为点不相交最大 S 圈分解问题(Vertex-Disjoint S -Cycle Packing, SVDCP)与边不相交最大 S 圈分解问题(Edge-Disjoint S -Cycle Packing, SEDCP)。一个图的周长(Girth)是指这个图中长度最小的圈的大小。若图 G 的周长为 g ,对于 CP 问题,如果限制集合 F 中所有圈的大小都为该固定值 g ,那么该类 CP 问题可描述为在图 G 中找最多个数的大小为 g 的不相交圈,即 g -最大圈分解问题(g -CP)。同理,按照点不相交与边不相交两种不同的分类方式,问题可以分为 g -点不相交最大圈分解问题(g -Vertex-Disjoint Cycle Packing, g -VDCP)与 g -边不相交最大圈分解问题(g -Edge-Disjoint Cycle Packing, g -EDCP)。下面给出这些问题的具体定义。

(1)有向图或无向图上的 VDCP(EDCP)问题

输入:图 $G=(V, E)$ (有向图或无向图),其中 V, E 分别为 G 的点集和边集。

输出: G 中两两之间没有公共顶点(边)的圈的个数的最大值。

(2)有向图或无向图上的 SVDCP(SEDCEP)问题

到稿日期:2015-11-17 返修日期:2016-03-18 本文受国家自然科学基金(61402054, 61502054, 61370172)资助。

罗卫东(1989-),男,硕士生,主要研究方向为参数计算;王建新(1969-),男,博士,教授,博士生导师,主要研究方向为计算机算法、网络优化理论、生物信息学;冯启龙(1982-),男,博士,副教授,主要研究方向为参数计算、计算机理论, E-mail: lyqilong@163.com(通信作者)。

输入:图 $G=(V,E)$ (有向图或无向图),其中 V,E 分别为 G 的点集和边集, S 为 V 的子集。

输出: G 中两两之间没有公共顶点(边)的 S 圈的个数的最大值。

(3)有向图或无向图上的 g -VDCP(g -SEDCP)问题

输入:周长为 g 的图 $G=(V,E)$ (有向图或无向图),其中 V,E 分别为 G 的点集和边集。

输出: G 中两两之间没有公共顶点(边)且长度为 g 的圈的个数的最大值。

本文第 2 节介绍了与 CP 问题相关的图论方面的一些研究现状,分别从该问题基于有向图与无向图两方面讨论;第 3 节主要从一般形式与特殊形式的 CP 问题两方面介绍它们的近似算法与不可近似性方面的研究现状;第 4 节讨论了 CP 问题在参数理论方面的研究现状;最后是总结和展望。

2 最大圈分解问题在图论方面的研究

本节主要介绍当图有某些特性时,图中不相交圈的个数满足怎样的结论。本节按照有向图与无向图两种图类,分两部分来叙述该问题的研究现状。第一部分介绍无向图中有关最大圈分解问题的性质,主要介绍了最大圈分解问题与反馈集问题的关系,以及最大圈分解与图的性质之间的关系,该部分的主线是介绍 Erdős-Pósa 定理^[1]与该定理的各种扩展形式的研究历程,以及图中顶点度的大小与不相交圈的个数的关系,即 Corrádi-Hajani 定理^[38]与该定理的各种扩展形式的研究历程;第二部分介绍有向图中圈分解问题的性质,该部分的主线是 Younger^[18]猜想与该猜想的各种扩展形式的研究历程。

2.1 无向图

自从 Erdős 等人^[1]提出与 CP 问题相关的 Erdős-Pósa 定理后,出现了大量关于 CP 问题图结构的研究。Erdős-Pósa 定理对于无向图上的 VDCP 问题给出了如下结论:对于任意的正整数 k ,存在一个函数 $f(k)$,使得任何一个无向图 G 要么包含 k 个点不相交圈,要么有一个大小为 $f(k)$ 的顶点反馈集(Feedback Vertex Set),并且 $f(k)=O(k\log k)$ 。随后出现了大量与该定理相关的研究。容易知道,图 G 的顶点反馈集的数目肯定大于图 G 中点不相交圈的数目,所以 $k \leq f(k)$ 。Erdős 给出的证明相当于给了 $f(k)$ 一个上限,即 $k \leq f(k) \leq O(k\log k)$ 。Bollobás 在其未发表的论文中证明了当 $k=2$ 时 $f(k)=3$,该定理说明,对于任意的一个图,该图要么有两个点不相交圈,要么删除 3 个点后该图一定为一个森林。Voss 等人^[17]证明了 $f(3)=6$,并且得到 $9 \leq f(4) \leq 12$ 。Erdős 等人^[2]通过改进文献^[1]中的方法,证明了一个更一般的情形,即 $c_1 k \log k \leq f(k) \leq c_2 k \log k$,其中 c_1 与 c_2 均为某个常数,即得到了 $f(k)$ 取值更好的上界与下界。

Kakimura 等人^[36]证明了对于任意的正整数 k ,存在一个函数 $f(k)$,对每一个无向图 G 与 G 中点集 $V(G)$ 的一个子集 S ,使得 G 要么包含 k 个点不相交的 S 圈,要么有一个大小为 $f(k)$ 的顶点反馈集破坏 G 中所有的 S 圈,并且得到 $f(k)=40k^2 \log_2 k$,文中通过应用 Geelen^[42]中关于最大不相交路径的一个结论,通过适当的改进后得到了该定理的证明。Ponte-

corvi 等人^[33]证明了对于任意的正整数 k ,存在一个函数 $f(k)$,使得对每一个无向图 G , G 要么包含 k 个点不相交的 S 圈,要么有一个大小为 $f(k)$ 的顶点反馈集破坏 G 中所有的 S 圈,并且得到 $f(k)=ck \log k$,文中构造出一个实例使得 $f(k)=\Omega(k \log k)$,因此得到 $f(k)=ck \log k$ 为最优解的结果;而且该性质直接得到了最大 S 圈分解的一个近似比为 $O(\log n)$ 的多项式时间近似算法。

设图 $G=(V,E)$,其中 $|V|=n$, V 中某个点 v 的邻居节点集用 $D(v)$ 表示,且 $d(v)=|D(v)|$, K_n 表示大小为 n 的团(clique)。Corrádi 等人^[38]得到了如果无向图 G 满足某些性质,则图中一定会有一定数目的点不相交圈,人们把他们的发现称为 Corrádi-Hajani 定理。Corrádi-Hajani 定理指当 k 为正整数,且 $n \geq 3k$ 时,若对 G 中任意一个点 u 都满足 $d(u) \geq 2k$ 时,则 G 包含 k 个点不相交圈。特别地,如果 $n=3k$,定理表明图 G 中一定有 k 个点不相交三角形。以下为对该定理的改进历程。

Justesen^[40]证明了当 k 为正整数,且 $n \geq 3k$ 时,若 G 中任意两个不相邻的点 u,v 都满足 $d(v)+d(u) \geq 4k$,则 G 包含 k 个点不相交圈。容易得知该结果为 Corrádi-Hajani 定理更一般的情形。Wang^[41]证明了当 k,s 为正整数,且 $n \geq sk, s \geq 3, n=qk+r(0 \leq r \leq k-1)$ 时,若 G 中任意一个点 u 都满足 $d(u) \geq (s-1)k$,则 G 包含 k 个点不相交圈 $c_1, c_2, \dots, c_k$,并且 $s \leq L(c_i) \leq q(1 \leq i \leq k-r), s \leq L(c_i) \leq q+1(k-r < i \leq k)$,其中 $L(c_i)$ 表示圈 c_i 的长度。Enomoto 等人^[22]与 Wang 等人^[23]分别独立证明了当 k 为不小于 2 的整数,且 $n \geq 3k$ 时,若 G 中任意两个不相邻的点 u,v 都满足 $d(v)+d(u) \geq 4k-1$,则 G 包含 k 个点不相交圈。该定理对 Justesen 提出的定理做了小的改进。Matsumura 等人证明当 k 为不小于 2 的整数,且 $n \geq 3k+2$ 时,若 G 中任意 3 个互不相邻的点 u,v,w 都满足 $d(v)+d(u)+d(w) \geq 6k-2$,则 G 包含 k 个点不相交圈。Egawa 等人^[25]证明当 k 为不小于 3 的整数, d 为某个整数且 $d \geq 4k-1, n \geq 3k$ 时,若 G 中任意两个不相邻的点 u,v 都满足 $d(v)+d(u) \geq d$,则 G 包含 k 个点不相交圈,并且这些圈至少覆盖 G 中 $\min\{d,n\}$ 个点。该定理在这方面首次给出加入限制的最大点不相交圈的结论。Kawarabayashi 等人^[26]证明当 d,n 为两个整数且 $d \geq 7, n \geq 6$ 时,若 G 中任意两个不相邻的点 u,v 都满足 $d(v)+d(u) \geq d$,则 G 包含 2 个点不相交圈并且这些圈至少覆盖 G 中 $\min\{d,n\}$ 个点。Brandt 等人^[27]证明当 k,n 为两个整数且 $n \geq 4k-1$ 时,若 G 中任意两个不相邻的点 u,v 都满足 $d(v)+d(u) \geq n$,则 G 包含 k 个点不相交圈,并且这些圈覆盖 G 中所有点。Enomoto 等人^[28]证明当 k,n 为两个整数且 $n \geq k$ 时,若 G 中任意两个不相邻的点 u,v 都满足 $d(v)+d(u) \geq n-k+1$,则除非 $k=2$ 并且 G 为一个长度为 5 的圈,否则 G 包含 k 个点不相交子图并且这些子图覆盖 G 中所有点,这些子图要么是圈,要么是孤立点或孤立边。Kawarabayashi 等人^[29]证明当 k,n 为两个整数且 $k \geq 2, n \geq 4k$ 时,若 G 中任意两个不相邻的点 u,v 都满足 $d(v)+d(u) \geq n-1$,则下列 3 条至少满足一条。

- 1) G 包含 k 个点不相交圈,并且这些圈覆盖 G 中所有点;
- 2) G 中删除某个大小为 $(n-1)/2$ 的点集后,剩下的点为原图 G 的独立集;

3) G 为一个 n 个点的图,该图由一个 $n-1$ 个点的团再加上一个点和该点与团上某个点的一条边构成。

Hu 等人^[30]证明当 k, n 为两个整数且 $n \geq 10k+3$ 时,若 G 中任意两个不相邻的点 u, v 都满足 $d(v)+d(u) \geq n-k+1$, 则 G 包含 k 个点不相交子图并且这些子图覆盖 G 中所有点,这些子图要么是圈,要么是孤立点。Fujita 等人^[31]证明当 k, n, r 为整数且 $n \geq 7k, 2 \leq r \leq k-2$ 时,若 G 中任意两个不相邻的点 u, v 都满足 $d(v)+d(u) \geq n-r$, 则 G 包含 k 个点不相交子图并且这些子图覆盖 G 中所有点,这些子图有 r 个是孤立点, $k-r$ 个是圈。Fujita 等人^[32]综合上述结论,得到一个推理,证明当 k, n, r 为整数且 $n \geq 10k+3, 0 \leq r \leq k-1, k \geq 2$ 时,若 G 中任意两个不相邻的点 u, v 都满足 $d(v)+d(u) \geq n-r$, 则 G 包含 k 个点不相交子图并且这些子图覆盖 G 中所有点,这些子图有 r 个是孤立点或圈, $k-r$ 个是圈。Chiba 等人^[39]证明当 k, r 为两个整数且 r 为某个常数, $n \geq r$ 时,若 G 中任意一个点 u 都满足 $d(u) \geq 2k$, 则以下两点必定满足一点。

1) G 包含 k 个点不相交的偶圈;

2) $(2k-2)K_1 + pK_2 \subseteq G \subseteq K_{2k-1} + pK_2 (2 \leq k \leq p)$, 或 $k=1$ 且 G 中的每一个块是一个 K_2 或一个偶圈。特别地, G 的每一个末端块都为偶圈。

2.2 有向图

Erdős-Pósa 定理讨论的是无向图中的情形。研究人员猜想在有向图中是否也有类似的定理存在,于是 Younger^[18]提出了如下猜想:对于每一个自然数 k , 都存在一个自然数 $f(k)$ ($g(k)$) 对于任何一个有向图 $G=(V, E)$ 都至少满足下列一条。

1) G 中有 k 个点(边)不相交的圈,这些圈组成集合 C ;

2) X 为 V 的子集(E 的子集), $|X| \leq f(k)$ ($|X| \leq g(k)$), 且 C 中所有的圈都至少有一个点(一条边)在 X 中。

人们把该猜想称为 Younger 猜想。

Gallai 证明了 $k=1$ 时猜想成立。McCuaig^[19]证明当 $k=2$ 时猜想成立,且 $f(2)=3$, 即对于任意的有向图 G , 要么 G 有两个点不相交的有向圈, 要么存在一个由 3 个点组成的顶点反馈集。Soares 证明如果 $f(k)$ 和 $g(k)$ 存在, 则 $f(k)$ 与 $g(k)$ 相等。Metzlar 等人^[43]证明如果限制在平面图 G 上, 并且 G 中每个点的最大出度与最大入度都不大于 2, Younger 猜想成立。特别地, 本文还得到了一个很好的顶点反馈集的上界, 即图 G 中要么含有 k 个点不相交的有向圈, 要么包含一个大小为 $4k$ 的顶点反馈集。Reed^[35]证明 Younger 猜想在平面向图上成立, 并给出平面图 G 中要么含有 k 个点不相交的有向圈, 要么包含一个大小为 $O(k \log k \log \log k)$ 的顶点反馈集。Seymour^[20]证明如果 $f(k)$ 存在, 则 $f(k) = \Omega(k \log k)$, 该猜想最终由 Reed 等人证明。Reed 等人^[21]巧妙地应用了 Ramsey 定理和网格方法证明了 Younger 猜想, 证明过程使用了归纳法, 由于 $k=1$ 时猜想成立已被证明, 因此通过对 k 进行规约, 假设 Younger 猜想在 k 时成立, 再证明 $k+1$ 时也成立, 从而证明了对任意整数 k , 猜想均成立。在归纳的过程中, 通过构造一种很大的网格得到有 k 个 VDCP 的图, 在构造的过程中使用了很多有向无环图的性质, 最典型的是, 若 T 是所有有

向圈的一个最小碰撞集 (Minimum Hitting Set), 那么 $G-T$ 为一个有向无环图。

随后人们考虑了更一般化的 Younger 猜想, 即要求有向图 G 得到的 k 个圈必须为 S 圈。Kakimura 等人^[37]证明了对于任意的正整数 k , 存在一个函数 $f(k)$, 对每一个有向图 G 与 G 中点集 $V(G)$ 的一个子集 S , 使得 G 要么包含 k 个 S 圈, 且 G 中的任何一个点至多同时在这 k 个圈的某 5 个圈中, 要么有一个大小为 $f(k)$ 的顶点反馈集。Kawarabayashi 等人^[34]证明了对于任意的正整数 k , 存在一个函数 $f(k)$, 对每一个有向图 G 与 G 中点集 $V(G)$ 的一个子集 S , 使得 G 要么包含 k 个 S 圈, 且 G 中的任何一个点至多同时在这 k 个圈的某两个圈中, 要么有一个大小为 $f(k)$ 的顶点反馈集。本文通过改进 Reed 等人^[21]提出的方法, 多次应用 Ramsey 定理, 得到了该结论。文中举出了一个反例, 说明在有向图 G 中不可能有如下性质, 即对于任意的正整数 k , 存在一个函数 $f(k)$, 对每一个有向图 G 与 G 中点集 $V(G)$ 的一个子集 S , 使得 G 要么包含 k 个点不相交的 S 圈, 要么有一个大小为 $f(k)$ 的顶点反馈集。构造该实例的思路同样是在 Reed 等人^[21]提出的构造大规模网格方法的基础上加入新的技术, 从而得到该结论, 该结论也证明了文献^[34]得到的关于一般化的 Younger 猜想的结论即为最优解, 从而证明了一般化的 Younger 猜想是不成立的。

3 最大圈分解问题在近似算法与不可近似性方面的研究

本节主要介绍 CP 问题的近似算法与该问题的不可近似性^[9]方面的研究, 分别从一般形式 CP 问题(即有向图与无向图上的 EDCP 问题与 VDCP 问题)和特殊形式的 CP 问题两个方面来介绍。

3.1 一般形式的 CP 问题

Holyer^[6]通过把 3-SAT 问题多项式规约到边分割问题, 证明了边分割问题是 NP-hard 的, 并在推论中得到一系列经典问题是 NP-hard 的, 其中包括 EDCP 问题, 而且可以很容易得到 VDCP 问题也为 NP-hard 的。由于 CP 问题是 NP-hard 的, 除非 $NP=P$, 否则 CP 问题没有多项式时间算法, 因此人们希望得到一个多项式时间的近似算法。对于近似算法, 如果求最大化问题, 问题的最优解除以近似算法在最坏情况下找到的解被定义为近似比; 如果求最小化问题, 近似比为近似算法在最坏情况下找到的解除以问题的最优解。近似比的大小是衡量近似算法好坏的一个重要标准, 近似比越接近 1, 则近似算法得到的解越接近最优解。近似比的值减去 1 称为误差值。若一个问题存在一个常数近似比的多项式时间近似算法, 则称该问题在 APX (Approximable) 中。若一个问题能在 $O(n^{\exp(1/\delta)})$ 的时间复杂度内得到一个 $1+\xi$ 近似比的近似算法, 则称该问题为 PTAS (Polynomial Time Approximation Scheme) 的。PTAS 规约是指一个问题 A 规约到一个问题 B 使用 3 个多项式时间可计算函数 f, g, h 并且满足以下条件:

1) 对于函数 f , 输入问题 A 中的任意一个实例, 输出为问题 B 中的某个实例;

2) 对于函数 g , 输入问题 A 的一个实例 x 、问题 B 的实例 $f(x)$ 的一个解和一个误差参数 ξ , 输出为实例 x 的一个近似解;

3) 对于函数 h , 输入为问题 A 的某个实例的误差参数 ξ , 输出为问题 B 的某个实例的误差参数 $h(\xi)$;

4) 如果 y 为实例 $f(x)$ 的近似解且近似比为 $1+h(\xi)$, 则实例 x 的解 $g(x, y, \xi)$ 的近似比为 $1+\xi$ 。

如果 APX 中的所有问题都可以通过 PTAS 规约规约到某个问题 A , 那么就称问题 A 为 APX-hard 的。如果一个问题为 APX-hard 的, 说明该问题多项式的时间近似算法的近似比不会小于某个常数 c , 除非 $P=NP$ 。

Caprara 等人^[3]最早给出了无向图上的 EDCP 问题的近似算法并证明了该问题是 APX-hard 的。他们通过简单的贪心方法给出了一个近似比为 $O(\log n)$ 的多项式时间近似算法, 但并没有证明该近似比是否为最佳近似比。文中还给出了一个实例说明贪心算法不可能做出某个多项式时间近似算法使得其近似比优于 $\Omega((\log n / (\log \log n))^{1/2})$ 。文中使用 PTAS 规约把 MAX-2-SAT 问题规约到 EDCP 问题, 由于 MAX-2-SAT 问题是 APX-hard 的, 因此证明了无向图的 EDCP 问题为 APX-hard 的。最后对问题做了线性规划松弛, 从整数间隙的角度再次简单地证明了 Erdős-Pósa 定理的正确性。

Krivelevich 等人^[7]通过改进文献^[3]中使用的贪心方法, 得到了无向图的 EDCP 问题的一个近似比为 $O((\log n)^{1/2})$ 的多项式时间近似算法。若要求无向图的 EDCP 问题中得到的圈都为 S 圈, 则可以得到一个近似比为 $O(n^{2/3})$ 的多项式时间近似算法。本文还对有向图的 EDCP 问题给出了一个近似比为 $O(n^{1/2})$ 的多项式时间近似算法, 并证明该问题不可能有近似比为 $O((\log n)^{1-\epsilon})$ 的多项式时间近似算法, 除非 $NP \subseteq ZPTIME(n^{\text{polylog} n})$ 。对于有向图的 VDCP 问题, 同样有近似比为 $O(n^{1/2})$ 的多项式时间近似算法。

Friggstad 等人^[8]证明了无向图的 EDCP 问题有一个大小为 $\Omega((\log n)^{1/2} / (\log \log n))$ 的整数间隙; 还证明了对于无向图的 EDCP 问题, 如果有近似比为 $O((\log n)^{1/2-\epsilon})$ 的多项式时间近似算法, 则会导致 $NP \subseteq ZPTIME(n^{\text{polylog} n})$, 该证明思路来源于对最大路径分解问题的不可近似性证明, 其证明方法是构造 3SAT 问题的实例, 得出如果无向图的 EDCP 问题有近似比为 $O((\log n)^{1/2-\epsilon})$ 的多项式时间近似算法, 则 3-SAT 在 $\text{co-RPTIME}(n^{\text{polylog} n})$ 中, 即 3SAT 有一个时间复杂度为 $ZPTIME(n^{\text{polylog} n})$ 的精确算法, 从而得到 $NP \subseteq ZPTIME(n^{\text{polylog} n})$ 。由于任意的点不相交圈一定是边不相交圈, 并且用类似方法证明了无向图上的 EDCP 问题的不可近似性, 得到无向图的 VDCP 问题同样不可能有近似比为 $O((\log n)^{1/2-\epsilon})$ 的多项式时间近似算法, 除非 $NP \subseteq ZPTIME(n^{\text{polylog} n})$ 。本文使用改进的贪心方法得到了无向图的 VDCP 问题的一个近似比为 $O(\log n)$ 的多项式时间近似算法。

综上所述, 对于无向图的 EDCP 问题, 人们已经找到了一个 $O((\log n)^{1/2})$ 的多项式时间近似算法, 并且证明该近似比很有可能是该问题的近似比下界。对于有向图的 EDCP 问题, 最好的多项式时间近似算法的近似比为 $O(n^{1/2})$, 近似比下界为 $O((\log n)^{1-\epsilon})$ 。对于无向图的 VDCP 问题, 最好的多项式时间近似算法的近似比为 $O(\log n)$, 近似比下界为 $O((\log n)^{1/2-\epsilon})$ 。对于有向图的 VDCP 问题, 最好的多项式时

间近似算法的近似比为 $O(n^{1/2})$ 。因此除了无向图的 EDCP 问题, 其余问题的近似算法或近似比下界还存在很大的改进空间。

3.2 特殊情形的 CP 问题

对于图 $G=(V, E)$, 其中 $n=|V|, m=|E|$, 若有 $m \geq \alpha n^{1+1/t+\epsilon}$, t 为正整数, α 与 ϵ 为任意固定的正实数, 则称该图为 $g(\alpha, t, \epsilon)$ 图。对于 $g(\alpha, t, \epsilon)$ 图上的 EDCP 问题, Holyer 等人^[6]证明了该问题是 NP-hard 的。Caprara 等人^[3]通过改进一般图上的 EDCP 问题的贪心算法, 给出了该问题的一个近似比可无限趋近于 $2t/3$ 的多项式时间近似算法, 但没有证明该问题是否为 APX-hard 的。

对于平面图上的 EDCP 问题, Caprara 等人^[3]证明了在平面图上即使所有点的度不大于 3, 该问题仍为 NP-hard 的。证明的方法是把平面 3-SAT 问题规约到该问题上, 从而得到上述结论。Caprara 等人^[4]发现该问题与独立集问题的一些独特关系, 通过独立集问题的近似算法得到了该问题一个近似比为 $2+\epsilon$ 的多项式时间近似算法。

对于平面 Eulerian 图上的 EDCP 问题, Frank 等人^[4,5]证明该问题是 P 的, 并给出了时间复杂度为 $O(mn)$ 的精确算法, 该证明方法利用了 Gallai-Edmonds 结构理论。

给定图 G , 对于 g -EDCP 问题, 当 $g=3$ 时, Garey 等人^[10]证明了该问题为 NP-hard 的。Caprara 等人^[11]证明了 G 中最大度大于 4 时, 该问题是 NP-hard 的, 最大度小于或等于 4 时该问题在 P 中。文中的证明方法为按图中顶点的最大度进行分类讨论, 在最大度为 3 的图中, 由于已知 3-VDCP 问题是 P 的, 因此很容易通过 3-VDCP 问题转换得到 3-EDCP 问题也在 P 中; 若最大度为 4, 通过构造 G 的一个辅助图, 把该问题转换成在辅助图上的最大独立集问题, 而该辅助图是一个不包含 claw 形状的诱导子图的图 (即 claw-free 图), 而在 claw-free 图上, 最大独立集问题是 P 的, 从而证明了最大度小于或等于 4 时 3-EDCP 在 P 中。Caprara 等人^[11]通过把 MAX-2-SAT-3 问题 PTAS 规约到 3-EDCP 问题上, 证明了最大度等于 5 时, 该问题是 APX-hard 的。在平面图上, 最大度等于 5 时, Caprara 等人^[11]通过把平面 3-SAT 多项式规约到该问题上, 证明了该问题是 NP-hard 的。

给定图 G , 对于 g -VDCP, 当 $g=3$ 时, Garey 等人^[6]证明了该问题为 NP-hard 的。Caprara 等人^[11]证明了 G 中最大度大于 3 时, 该问题是 NP-hard 的。最大度小于或等于 3 时该问题在 P 中, 证明时把图分为两类: 一类是图中顶点最大度为 2, 这种情况下该问题在 P 中是显然的; 另一类是图中顶点最大度为 3, 对于这类, G 中任意 4 个点要么为 4 的团, 要么为钻石结构, 通过处理这两种结构得到该种情况也在 P 中。综合以上两种情况, 最终得到在最大度小于 4 的图上, 该问题在 P 中。

Kann 等人^[44]证明了 3-VDCP 问题是 APX-hard 的。图 G 中最大度等于 4 时, Caprara 等人^[11]证明该问题是 APX-hard 的, 文中通过把 MAX-2-SAT-3 问题 PTAS 规约到该问题上, 证明了该结论。Hurkens 等人^[46]得到了该问题的一个近似比为 $1.5+\epsilon$ 的多项式时间近似算法。

对于平面图上的 3-VDCP 问题, 最大度等于 4 时, Caprara 等人^[11]证明了该问题是 NP-hard 的, 文中通过把平面

3-SAT 多项式规约到该问题上证明了它的难度。

Rautenbach 等人^[12]证明了 $g=4,5$ 并且 G 中最大度大于或等于 4 时, g -EDCP 与 g -VDCP 问题为 APX-hard 的,若 G 中最大度小于 4,则该问题在 P 中。对于 $g \geq 6$ 并且 G 中最大度大于或等于 3 时, g -EDCP 与 g -VDCP 问题为 APX-hard 的,若 G 中最大度小于 3,则该问题在 P 中。文中首先证明了一个中间问题 3-OCC MAX 2-SAT 是 APX-hard 的,然后将该问题 PTAS 规约到上述问题中,从而证明了上述结论。

4 最大圈分解问题在参数计算方面的研究

在实际应用中,近似算法得到的解并非在所有情形下都能得到很好的应用。近些年,随着参数计算理论^[19-23]发展得越来越成熟,许多使用精确算法与近似算法得不到好的结果的问题在使用参数算法时能得到满意的结果。参数计算成了处理 NP 难问题的一个新思路,它把理论和实践很好地结合到了一起。一个问题参数化版本的输入由两部分构成,一部分是问题的某个实例,另一部分是该实例对应的参数 k 。如果处理该问题的算法的时间复杂度为 $f(k)n^{O(1)}$,即该算法的时间复杂度的指数部分只与 k 有关而与实例的长度 n 无关,就称该问题是固定参数可解的(Fixed-Parameter Tractable, FPT)。当且仅当问题可以核心化时一个问题为 FPT 的。参数计算理论中,对于参数化问题,如果按问题的参数复杂性分类,则可分为 FPT 和 W 谱系,其中 W 谱系又可分为 $W[1], W[2], W[3], \dots$,人们一般认为 FPT 与 $W[i] (i=1,2,3, \dots)$ 两两之间互不相等,而且一个问题只有在 FPT 中才说这个问题是有效参数算法的。对参数计算理论的进一步了解可参考文献[24,52,53]。

参数化的 VDCP, EDCP 与 SVD, SEDCP 问题的定义如下。

(1) 有向图或无向图上的 VDCP (EDCP) 问题的参数形式。

输入: 图 $G=(V, E)$ (有向图或无向图), 其中 V, E 分别为 G 的点集和边集。

参数: k 。

输出: G 中是否有 k 个点(边)不相交的圈。

(2) 有向图或无向图上的 SVD (SEDCP) 问题的参数形式。

输入: 图 $G=(V, E)$ (有向图或无向图), 其中 V, E 分别为 G 的点集和边集, S 为 V 的子集。

参数: k 。

输出: G 中是否有 k 个点(边)不相交的 S 圈。

(3) 有向图或无向图上的 g -VDCP (g -SEDCP) 问题的参数形式。

输入: 周长为 g 的图 $G=(V, E)$ (有向图或无向图), 其中 V, E 分别为 G 的点集和边集。

参数: k 。

输出: G 中是否有 k 个点(边)不相交的周长为 g 的圈。

对于有向图上参数化的 VDCP 问题, Grohe 等人^[14]参考 Slivkins^[13]证明最大边不相交路径问题(Edge-Disjoint Paths)为 $W[1]$ 完全的方法,证明了参数化的 VDCP 问题是 $W[1]$ 完全的,即该问题没有有效参数算法,除非 $FPT=W[1]$ 。Gro-

he 等人^[14]对参数化的 VDCP 问题做了第一个 FPT 近似算法,即如果图 G 中最大点不相交圈的个数为 k ,那么可以在 FPT 时间内得到一个 $k/p(k)$ 的解,并且函数 $k/p(k)$ 关于 k 是一个非递减的函数。该问题使用的方法是以文献[21]的结论为基础的。

对于无向图上的 VDCP 问题, Bodlaender 等人^[15]证明了该问题没有多项式核,除非 PH (Polynomial Hierarchy) 垮到第三层。文中首先给出了一个被称为不相交因子(Disjoint Factors)的中间问题,并证明了该问题满足蒸馏算法(Distillation Algorithm)的性质,从而得到不相交因子问题没有多项式核的结论,再把该问题规约到无向图上的参数化的 VDCP 问题上,从而证明了该问题没有多项式核。Bodlaender 等人^[16]证明该问题如果限制在平面图上有线性核,对于 3-EDCP 问题的参数形式,有一个大小为 $4k$ 的核。

对于有向图上的 SVD 问题的参数形式, Kakimura 等人^[37]得到了该问题的一个 FPT 近似算法,即如果图中存在 k 个点不相交的 S 圈,则算法在 $f(k)n^{O(1)}$ 的时间内可以得到一个 $k/p(k)$ 的解,其中 $p(k)$ 为 k 的函数,并且 $k/p(k)$ 是一个关于 k 的非递减函数。Kawarabayashi 等人^[34]在解决一般化的 Younger 猜想的过程中,对有向图上参数化的 SVD 问题有了更深一步的观察,从而得到了该问题的一个更好的 FPT 近似算法。

结束语 CP 问题是一类重要的算法与图论问题,它在很多领域都有重要的应用。本文阐述了 CP 问题的研究现状,可以看出,该问题的主要研究集中在图论方面,而算法方面的研究相对较少。人们在对该问题研究的过程中得到了丰富的结论,并且更深一步地认识了 CP 问题,但很多方面还可做进一步的研究,可主要从以下方面出发。

图论方面的大量研究加深了人们对 CP 问题的认知,文章提到的各类图论方面的结果仍可继续进行优化,如 Corrádi-Hajnal 定理是否能够改进到更好的形式?

近似算法与不可近似性方面,对于有向图的 EDCP 问题,最好的多项式时间近似算法的近似比为 $O(n^{1/2})$,近似比下界为 $O((\log n)^{1-\epsilon})$ 。对于无向图的 VDCP 问题,最好的多项式时间近似算法的近似比为 $O(\log n)$,近似比下界为 $O((\log n)^{1/2-\epsilon})$ 。对于有向图的 VDCP 问题,最好的多项式时间近似算法的近似比为 $O(n^{1/2})$,没有相关的近似下界。理论上这些问题的近似算法或者近似比下界还有很大的改进空间。许多特殊情形的 CP 问题仅得到了它们的复杂度,而没有相关的近似算法与不可近似性的研究,这方面大量的工作需要完善。

该问题参数算法的研究结论还不多,对于有向图 VDCP 问题,可以对其进行改进以得到更好的参数近似算法,可以考虑平面图上的 VDCP 问题是不是 FPT 的。对无向图中的 VDCP 问题可以设计参数算法。特殊形式的 CP 问题在参数理论方面的研究很少,这些问题在参数领域中的未知结论都可以成为研究方向。

据作者所知,各类 CP 问题没有精确算法方面的研究,得到各类 CP 问题一些好的精确算法,在理论与实践上都是有意义的。

参考文献

- [1] ERDŐS P, PÓSA L. On the maximal number of disjoint circuits of a graph[J]. Publicationes Mathematicae Debrecen, 1962(9): 3-12.
- [2] ERDŐS P, PÓSA L. On independent circuits contained in a graph[J]. Canada Journal of Mathematics, 1965, 17: 347-352.
- [3] CARPARA A, PANCONESI A, RIZZI R. Packing cycles in undirected graphs[J]. Journal of Algorithms, 2003, 48(1): 239-256.
- [4] CARPARA A, PANCONESI A, RIZZI R. Packing cuts in undirected graphs[J]. Networks, 2004, 44(1): 1-11.
- [5] FRANK A. Conservative weightings and ear decompositions of graphs[J]. Combinatorica, 1993, 13(1): 65-81.
- [6] HOLYER I. The NP-completeness of some edge-partition problems[J]. SIAM Journal on Computing, 1981, 10: 713-717.
- [7] KRIVELEVICH M, NUTOV Z, YUSTER R. Approximation algorithms for cycle packing problems[C]// Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms. Vancouver, British Columbia, Canada, 2005: 556-561.
- [8] FRIGGSTAD Z, SALAVATIPOUR M R. Approximability of Packing Disjoint Cycles[J]. Algorithmica, 2010, 60(2): 395-400.
- [9] DU Ding-zhu, GE Ke-yi, HU Xiao-dong. Design and analysis of approximation algorithms[M]. Beijing: China Higher Education Press, 2011. (in Chinese)
堵丁柱, 葛可一, 胡晓东. 近似算法的设计与分析[M]. 北京: 高等教育出版社, 2011.
- [10] GAREY M R, JOHNSON D S. Computers and Intractability: A Guide to the Theory of Incompleteness[M]. San Francisco, 1979.
- [11] CAPRARA A, RIZZI R. Packing triangles in bounded degree graphs[J]. Information Processing Letters, 2002, 84(4): 175-180.
- [12] RAUTENBACH D, REGEN F. On packing shortest cycles in graphs[J]. Information Processing Letters, 2009, 109: 816-821.
- [13] SLIVKINS A. Parameterized tractability of edge disjoint paths on directed acyclic graphs[J]. SIAM Journal on Discrete Mathematics, 2010, 24(1): 146-157.
- [14] GROHE M, GRUBER M. Parameterized Approximability of the Disjoint Cycle Problem[C]// Automata, Languages and Programming, 34th International Colloquium. Wroclaw, Poland, 2007: 363-374.
- [15] BODLAENDER H L, JANSEN B M P, KRATSCHE S. Kernel bounds for path and cycle problems[J]. Theoretical Computer Science, 2013, 511: 117-136.
- [16] BODLAENDER H L, PENNINKX E, TAN R B. A Linear Kernel for the k -Disjoint Cycle Problem on Planar Graphs[C]// Algorithms and Computation, 19th International Symposium. Gold Coast, Australia, 2008: 306-317.
- [17] VOSS H J, JÜRGEN H, GRAPHERN E V. die keine $k+1$ knotenferne Kreise enthalten[J]. Mathematische Nachrichten, 1969, 40: 19-25.
- [18] YOUNGER D H. Graphs with interlinked directed circuits[C]// Proceedings of the International Midwest Symposium on Circuit and Systems. 1973: XVI 2. 1-XVI 2. 7.
- [19] MCCUAIG W. Intercyclic digraphs. Graph Structure Theory [M]// AMS Contemporary Math, 1991: 203-245.
- [20] SEYMOUR P D. Packing Directed Circuits Fractionally [J]. Combinatorica, 1995, 15(2): 281-288.
- [21] REED B, ROBERTSON N, Seymour P, et al. Packing Directed Circuits[J]. Combinatorica, 1996, 16(4): 535-554.
- [22] ENOMOTO H. On the existence of disjoint cycles in a graph [J]. Combinatorica, 1998, 18(4): 487-492.
- [23] WANG Hong. On the maximum number of independent cycles in a graph[J]. Discrete Mathematics, 1999, 205(1-3): 183-190.
- [24] 王建新, 冯启龙. 参数计算导论[M]. 长沙: 科学出版社, 2014.
- [25] EGAWA Y, HAGITA M, KAWARABAYASHI K, et al. Covering vertices of a graph by k disjoint cycles[J]. Discrete Mathematics, 2003, 270(1-3): 114-124.
- [26] EGAWA Y, FUJITA S, KAWARABAYASHI K, et al. Existence of two disjoint long cycles in graphs[J]. Discrete Mathematics, 2005, 305(1-3): 154-169.
- [27] BRANDT S, CHEN G, FAUDREE R, et al. Degree conditions for 2-factors[J]. Journal of Graph Theory, 1997, 24(2): 165-173.
- [28] ENOMOTO H, LI H. Partition of a graph into cycles and degenerated cycles[J]. Discrete Mathematics, 2004, 276(1-3): 177-181.
- [29] KAWARABAYASHI K. Degree sum conditions and graphs which are not covered by k cycles[J]. Discrete Mathematics, 2008, 308: 1620-1627.
- [30] HU Zhi-quan, LI Hao. Weak cycle partition involving degree sum conditions[J]. Discrete Mathematics, 2009, 309(4): 647-654.
- [31] FUJITA S. Degree conditions for the partition of a graph into cycles, edges and isolated vertices[J]. Discrete Mathematics, 2009, 309(11): 3534-3540.
- [32] FUJITA S. Recent Results on Disjoint Cycles in Graphs[J]. Electronic Notes in Discrete Mathematics, 2005, 22: 409-412.
- [33] PONTECORVI M, WOLLAN P. Disjoint cycles intersecting a set of vertices[J]. Journal of Combinatorial Theory, Series B, 2012, 102(5): 1134-1141.
- [34] KAWARABAYASHI K, KRÁL D, KRČÁL M, et al. Packing directed cycles through a specified vertex set[C]// Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms. New Orleans, Louisiana, USA, 2013: 365-377.
- [35] REED B, SHEPHERD B. The Gallai-Younger conjecture for planar graphs[J]. Combinatorica, 1996, 16: 555-566.
- [36] KAKIMURA N, KAWARABAYASHI K, MARX D. Packing cycles through prescribed vertices[J]. Journal of Combinatorial Theory, 2011, 101(5): 378-381.
- [37] KAKIMURA N, KAWARABAYASHI K. Packing Directed Circuits through Prescribed Vertices Bounded Fractionally[J]. SIAM Journal of Discrete Mathematics, 2012, 26(3): 1121-1133.
- [38] CORRÁDI K, HAJNAL A. On the maximal number of independent circuits in a graph[J]. Acta Mathematica, 1963, 14: 423-443.
- [39] CHIBA S, FUJITA S, KAWARABAYASHI K, et al. Disjoint Even Cycles Packing[J]. Electronic Notes in Discrete Mathematics, 2009, 34: 113-119.

- [16] LIN Y, LIU T, CHEN H. Semantic manifold learning for image retrieval[C]// ACM International Conference on Multimedia. 2005; 249-258.
- [17] YU J, TIAN Q. Learning image manifolds by semantic subspace projection[C]// ACM International Conference on Multimedia. 2006; 297-306.
- [18] MIAO D, GAO C, ZHANG N, et al. Diverse reduct subspaces based co-training for partially labeled data[J]. International Journal of Approximate Reasoning, 2011, 52(8): 1103-1117.
- [19] SKOWRON A, RAUSZER C. The discernibility matrices and functions in information systems[M]// Slowinski R, ed. Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory. Dordrecht, Kluwer, 1992.
- [20] HU X H, CERONE N. Learning in relational databases: A rough set approach[J]. International Journal of Computational Intelligence, 1995, 11(2): 323-338.
- [21] YANG Ming. An incremental updating algorithm for attribute reduction based on improved discernibility matrix[J]. Chinese Journal of Computers, 2007, 30(5): 815-822. (in Chinese)
杨明. 一种基于改进差别矩阵的属性约简增量式更新算法[J]. 计算机学报, 2007, 30(5): 815-822.
- [22] BLUM A, MITCHELL T. Combining labeled and unlabeled data with co-training[C]// Proc of the 11th Annual Conf on Computational Learning Theory. New York: ACM, 1998; 92-100.
- [23] FEGER F, KOPRINSKA I. Co-training using RBF nets and different feature splits[C]// Proc of the 2006 Int Joint Conf on Neural Networks. Piscataway, NJ: IEEE, 2006; 1878-1885.
- [24] YASLAN Y, CATALTEPE Z. Co-training with relevant random subspaces[J]. Neurocomputing, 2010, 73(10-12): 1652-1661.
- [25] TANG Huan-ling, LIN Zheng-kui, LU Ming-yu, et al. An advanced co-training algorithm based on mutual independence and diversity measures[J]. Journal of Computer Research and Development, 2008, 45(11): 1874-1881. (in Chinese)
唐焕玲, 林正奎, 鲁明羽, 等. 一种结合独立性模型与差异评估的 Co-Training 改进方案[J]. 计算机研究与发展, 2008, 45(11): 1874-1881.
- [26] SALAHELDIN A, GAYAR N El. New feature splitting criteria for cotraining using genetic algorithm optimization[C]// Proc of the 9th Int Workshop on Multiple classifier systems. Berlin: Springer, 2010; 22-32.
- [27] GOLDMAN S, ZHOU Yan. Enhancing supervised learning with unlabeled data[C]// Proc of the 17th Int Conf on Machine Learning. San Francisco: Morgan Kaufmann, 2000; 327-334.
- [28] ZHOU Zhi-hua, LI Ming. Tri-training: Exploiting unlabeled data using three classifiers[J]. IEEE Trans on Knowledge and Data Engineering, 2005, 17(11): 1529-1541.
- [29] BLAKE C, MERZ C J. UCI Repository of machine learning databases[OL]. <http://archive.ics.uci.edu/ml/datasets.html>.
- [30] ØHRN A, KOMOROWSKI J. ROSETTA: A rough set toolkit for analysis of data[C]// Proceedings of the 3rd International Joint Conference on Information Sciences. Durham, NC, USA: 1997; 403-407.
- (上接第6页)
- [40] JUSTESEN P. On independent circuits in nite graphs and a conjecture of Erdős and Pósa[J]. Annals of Discrete Mathematics, 1989, 41: 299-306.
- [41] WANG Hong. Independent cycles with limited size in a graph[J]. Graphs and Combinatorial, 1994, 10: 271-281.
- [42] GELEN J, GERARDS B, REED B, et al. On the odd-minor variant of Hadwiger's conjecture[J]. Journal of Combinatorial Theory, Series B, 2009, 99: 20-29.
- [43] METZLAR A, MURTY U S R. Disjoint Circuits in Planar Digraphs[J]. Journal of Combinatorial, Theory Series B, 1993, 57(2): 228-238.
- [44] KANN V. Maximum bounded 3-dimensional matching is MAX SNP-complete[J]. Information Processing Letters, 1991, 37: 27-35.
- [45] AUSIELLO G, CRESCENZI P, Gambosi G, et al. Combinatorial Optimization Problems and Their Approximability Properties [M]. Berlin: Springer, 1999.
- [46] HURKENS C A J, SCHRIJVER A. On the size of systems of sets every t, of which have an SDR, with an application to the worstcase ratio of heuristics for packing problems[J]. SIAM Journal on Discrete Mathematics, 1989, 2(1): 68-72.
- [47] EIMABROUK N. Genome Rearrangement by Reversals and Insertions/Deletions of Contiguous Segments[C]// Combinatorial Pattern Matching, 11th Annual Symposium. Montreal, Canada, 2000; 222-234.
- [48] BAFNA V, PEVZNER P A. Genome Rearrangements and Sorting by Reversals[J]. SIAM Journal on Computing, 1996, 25(2): 272-289.
- [49] SHAO Ming-fu, LIN Yu. Approximating the edit distance for genomes with duplicate genes under DCJ, insertion and deletion [J]. BMC Bioinformatics, 2012, 13(S-19): S13.
- [50] SHAO Ming-fu, LIN Yu, BERNARD M. An Exact Algorithm to Compute the DCJ Distance for Genomes with Duplicate Genes [C]// Research in Computational Molecular Biology 19th Annual International Conference. Pittsburgh, PA, USA, 2014; 280-292.
- [51] CHEN Xin, ZHENG Jie, FU Zheng, et al. Assignment of orthologous genes via genome rearrangement[J]. IEEE/ACM Transactions on Computational Biology Bioinformatics, 2005, 2(4): 302-315.
- [52] LI Shao-hua, WANG Jian-xin, FENG Qi-long, et al. Kernelization techniques and its application to parameterized computation [J]. Journal of Software, 2009, 20(9): 2307-2319. (in Chinese)
李绍华, 王建新, 冯启龙, 等. 参数计算中核心化计算及其应用的研究进展[J]. 软件学报, 2009, 20(9): 2307-2319.
- [53] ZHOU Xing, PENG Wei. Several techniques used in parameterized computation[J]. Computer Science, 2014, 41(6A): 18-23. (in Chinese)
周星, 彭伟. 参数计算中使用的若干技术[J]. 计算机科学, 2014, 41(6A): 18-23.
- [54] LU Sheng-guo, LI Shuang, ZHU Jian-guo, et al. Survey of genome rearrangements[J]. China Biotechnology, 2013, 30(7): 108-111. (in Chinese)
卢胜国, 李霜, 朱建国, 等. 基因组重排技术应用及进展[J]. 中国生物工程杂志, 2013, 30(7): 108-111.