第 3 卷第 4 期
1992 年 11 月

软 件 学 报
JOURNAL OF SOFTWARE

Vol. 3, No. 4
Nov., 1992

图(k,m)最优划分的近似算法

吕其诚

(黑龙江大学计算机科学系,哈尔滨 150080)

AN APPROXIMATION ALGORITHM ON GRAPH (K,M) OPTIMAL PARTITION PROBLEM

Lü Qicheng

(*Department of Computer Science, Heilongjiang University, Harbin* 150080)

Abstract In this paper we present an approximation algorithm for (k,m) optimal partition problem on an undirected graph. We show that this approximation algorithm which produces a near optimal solution runs in polynomial time. The performance ratio in the worst case is bounded by a parameter k, which is independent of input size of the problem.

摘要 本文提出了无向图(k,m)最优划分的一个近似算法,证明了这是一个产生近似最优解的多项式时间算法.在最坏情况下,该算法的性能保证为一个参数 k 所界定,这里 k 是与问题输入尺寸无关的.

§1. 问题的概述

1.1 问题的定义

给定一个无向单图 $G=(V,E)$,其中结点集 V 中每一个结点皆有重量,边集 E 中每一条边皆有耗费(或权).另外图无结点自回路.一般所谓图最优划分问题,是把图的结点集 V 划分成若干互不相交的结点子集(相应的子图可保连通或不保连通),其中每一个子集中结点重量之和不超过给定的一个重量,且使该划分所形成的各子图间的外部边(即子集与子集间结点外部连线)耗费之和为最小.当子集数 $m\geqslant 3$ 时,该问题已经被证明是 NP—完全的[2].易知当各结点重量及各边耗费均是 1,各子集中结点数不超过一个给定常数 $k\geqslant 3$ 时,该问题仍是 NP—完全的.本文所讨论的正是这种条件下的图(k,m)最优划分问题.下面给出问题的定义.

图(k,m)最优划分问题——

实例条件:一个无结点自回路的无向单图 $G=(V,E)$,其中结点集 V 中有几个结点,依次编号为 $1,2,3,\cdots,n$,且每个结点重量为 1;边集 E 中有 $|E|=e$ 条边,各边耗费组成耗费矩阵 C

本文 1990 年 6 月 11 日收到,1990 年 11 月 16 日定稿.作者吕其诚,讲师,人工智能室副主任,主攻算法理论及应用,其中包括近似算法,并行算法及人工智能中的算法.

© 中国科学院软件研究所 http://www.jos.org.cn

$=(c_{ij})_{n\times n}$，满足 $c_{ij}=c_{ji}(i\neq j)$，$c_{ij}=0$ 当且仅当 $(i=j)$ 或 $(i,j)\notin E$；另外，有一个整数 $k\geqslant 3$.

问题：把结点集 V 划分成 m 个互不相交的子集 $V_1,V_2,\cdots,V_m$，这里 $m=\lceil n/k\rceil$，使得 $|V_i|=k(1\leqslant i\leqslant m-1)$，$|V_m|=n-(m-1)\cdot k$，且使该划分所有外部边耗费之和 $\sum_{(i,j)\in T}c_{ij}$ 为最小，这里 T 是外部边集合.

1.2 关于近似算法

众所周知，$NP=?P$ 的问题至今仍是令人注目的悬案. 七十年代以来证明 NP—完全性的一系列结果表明，除非 $NP=P$，否则出现在商业、科学以及工程中的大多数组合优化问题都是难解的，换言之，对这类问题我们还找不到多项式时间的有效算法. 在实际应用中我们应如何求解这些问题？许多学者认为，在很多场合下，寻求保证对这类问题产生近似最优解的多项式时间算法是可行的. 这种算法称为近似算法(approximation algorithm).

近似算法一般有两条标准：

(1)时间标准：其时间复杂度是关于 n 的多项式，这里 n 是问题的尺寸.

(2)精度标准：其性能比，即问题的最优解与算法生成的近似解的比值上有界.

现在已有不少关于图的最优划分算法，例如利用最大流最小割集方法，线性规划方法等. 这些算法中有不少性能差，或不能保证性能比上有界，或不能保证是多项式时间的有效算法. 本文给出的近似算法可以得出近似最优解，其性能比为一个与输入大小无关的常数 k 所界定，且证明了它是多项式时间的有效算法.

§2. 算法的描述

设 C_E 为图 $G=(V,E)$ 所有边耗费之和；C_T 是划分所产生的所有外部边耗费之和；C_I 是该划分所产生所有内部边耗费之和. 显然，对于图的一个 (k,m) 划分，任一边 $e\in E$，如果它不是外部边，就一定是内部边，反之亦然. 对于给定的图 G，其 C_E 值是确定的. 要使 $C_T=\sum_{(i,j)\in T}c_{ij}$ 的值最小，等价地，就是要使 C_I 的值最大. 这可由 $C_E=C_T+C_I$，即 $C_T=C_E-C_I$ 而知.

我们的算法采用了贪心法(gready algorithm)思想，其方法是：要构成结点划分 $H=\{V_1,V_2,\cdots,V_m\}$，可分 m 步进行，每一步选取一个结点子集. 在第 i 步选取子集 $V_i(1\leqslant i\leqslant m-1)$ 时依据原则是在 $V-\bigcup_{1\leqslant j\leqslant i-1}V_j$ 中，选取其结点所有 k—组合中结点所关联内部边耗费和最大者. 注意每一个 $|V_i|=k,i=1,2,\cdots,m-1$. 最后一步即第 m 步选取 $V_m=V-\bigcup_{1\leqslant j\leqslant m-1}V_j$.

依上所述，我们给出图 (k,m) 最优划分的近似算法(即 OGP)如下. 其问题定义如前所述.

```
Procedure OGP(G)
V'←V
m←⌈|V|/k⌉
for i=1 to m−1 do
    begin
    (i)从 V'中的所有 k−子集中选取所关联内部边耗费和最大者作为 Vi.
    (ii)V'←V'−Vi
    end
Vm←V'
end {procedure}
```

© 中国科学院软件研究所 http://www.jos.org.cn

§3. 算法的性能分析

3.1 算法的精度性能保证

我们先引入如下记号，以便于进行算法分析：

1. 算法 OGP 产生的结点划分是 $H=\{V_1,V_2,\cdots,V_m\}$.

2. 设 $A=\{A_1,A_2,\cdots,A_m\}$ 是任何一个 (k,m) 划分算法 A^* 所产生的结点划分；其中最优的 (k,m) 划分的结点划分记为 $O=\{O_1,O_2,\cdots,O_m\}$.

3. 设 S 是图 $G=(V,E)$ 的任意子图，记 $W(S)$ 为 S 中所有边(即内部边)耗费和. 对于图 G 的任意结点划分 $A=\{A_1,A_2,\cdots,A_m\}$，相应子图记为 $G_i=(A_i,E_i)$，这里 A_i 是 G_i 的结点集，$E_i=\{(p,q)\}|p,q\in A_i$ 且 $(p,q)\in E\}$ 是 G_i 的边集. 子图 G_i 的所有耗费和应记为 $W(G_i)$，为方便记为 $W(G_i)=W(A_i)$. 于是算法 A^* 所产生的划分其内部边耗费之和记为 $W(A)$，则我们有 $W(A)=\sum_{i=1}^{m}W(A_i)=\sum_{i=1}^{m}W(G_i)$.

4. 定义集合 $F_j(1\leqslant j\leqslant m)$ 如下：

$F_1=\{A_i|A_i\cap V_1\neq\varnothing,1\leqslant i\leqslant m\}$，

当 $2\leqslant j\leqslant m$ 时有

$F_j=\{A_i|A_i\cap V_j\neq\varnothing,A_i\notin F_1\cup F_2\cup\cdots\cup F_{j-1},1\leqslant i\leqslant m\}$

然后我们有如下的事实：

事实 1：$F_i\cap F_j=\varnothing$，对于 $1\leqslant i\neq j\leqslant m$；

事实 2：对任何 $A_i\in F_j$，有 $A_i\cap(V_1\cup V_2\cup\cdots\cup V_{j-1})=\varnothing$；

事实 3：$\bigcup_{1\leqslant j\leqslant m}F_j=A$.

我们还容易证明下面的引理和定理：

引理 1：对于实例条件中的整数 $k\geqslant 3$，有 $|F_j|\leqslant k$，$1\leqslant j\leqslant m$.

证明：对于每一个 $V_j\in H$，皆有 $|V_j|=k(1\leqslant j\leqslant m-1)$，或 $|V_m|\leqslant k$. F_j 是那些与 V_j 有公共结点的 A_i 的集合，由 $|V_j|\leqslant k(1\leqslant j\leqslant m)$ 知，这样的 A_i 至多有 k 个，从而有 $|F_j|\leqslant k$ $(1\leqslant j\leqslant m)$.

引理 2：如果 $A_i\in F_j$，则 $W(A_i)\leqslant W(V_j)$，$(1\leqslant i,j\leqslant m)$.

证明：对于 j 用数学归纳法.

当 $j=1$，必有 $W(A_i)\leqslant W(V_1)$，因为 V_1 是结点集 V 中所有结点 k—组合的内部边耗费和最大者.

假设当 $j=2,3,\cdots,r$ 时引理为真. 那么当 $j=r+1$ 时，如果有 $A_i\in F_{r+1}$，则 $A_i\notin F_1\cup F_2\cup\cdots\cup F_r$，由事实 2 已知 $A_i\cap(V_1\cup V_2\cup\cdots\cup V_r)=\varnothing$. 因此当由算法 OGP 产生 V_{r+1} 时，它必是集合 $V-\bigcup_{1\leqslant i\leqslant r}V_i$ 中所有 k—组合(其中有 A_i)的内部边耗费和最大者，故有 $W(A_i)\leqslant W(V_{r+1})$. 从而引理得证.

定理 1：$W(A)/W(H)\leqslant k$.

证明：对于图 G 的任意一个 (k,m) 划分算法 A^* 产生的结点划分是 $A=\{A_1,A_2,\cdots,A_m\}$，则 $W(A)=\sum_{i=1}^{m}W(A_i)$. 依事实 3，我们有 $F_1\cup F_2\cup\cdots\cup F_m=A$. 记 $W(F_j)=\sum_{A_i\in F_j}W(A_i)$，$(1\leqslant j$

© 中国科学院软件研究所 http://www.jos.org.cn

$\leqslant m$)，于是 $W(A)=\sum_{j=1}^{m} W(F_j)=\sum_{j=1}^{m}\sum_{A_i\in F_j} W(A_i)$.

由引理 2，有 $W(A_i)\leqslant W(V_j)$（当 $A_i\in F_j$），再由引理 1 知 $|F_j|\leqslant k$，即使 $A_i\in F_j$ 的 A_i 不超过 k 个，故 $W(F_j)=\sum_{A_i\in F_j} W(A_i)\leqslant k\cdot W(V_j)$，从而 $W(A)=\sum_{j=1}^{m} W(F_j)\leqslant k\cdot\sum_{j=1}^{m} W(V_j)=k\cdot W(H)$，即 $W(A)/W(H)\leqslant k$，证毕.

最优的 (k,m) 划分 $O=\{O_1,O_2,\cdots,O_m\}$ 也是一个 (k,m) 划分，故有 $W(O)/W(H)\leqslant k$. 这样，我们就给出了最优 (k,m) 划分与算法 OGP 产生的划分的内部边耗费和之比（即性能比）的上界为 k.

3.2 算法的最坏情况分析

通过对最坏情况研究和分析，可以知道我们给出的性能比上界 k 是最小上界. 我们有如下的定理：

定理 2：对于任何正整数 $k\geqslant 3$ 和任意正整数 $\varepsilon>0$，都可构造一个图 $G(k,\varepsilon)=(V(k,\varepsilon),E(k,\varepsilon))$，对于该图我们可以找到一个 (k,m) 划分 Q，使得 $W(Q)\geqslant(k-\varepsilon)\cdot W(H)$.

证明：依下面方法可构造一个具有 k^2 个结点和 k^2-1 条边的无向单图 $G(k,\varepsilon)=(V(k,\varepsilon),E(k,\varepsilon))$，使得

结点集 $V(k,\varepsilon)=\{v_{pq}\mid p=1,2,\cdots,k$ 和 $q=1,2,\cdots,k\}$，

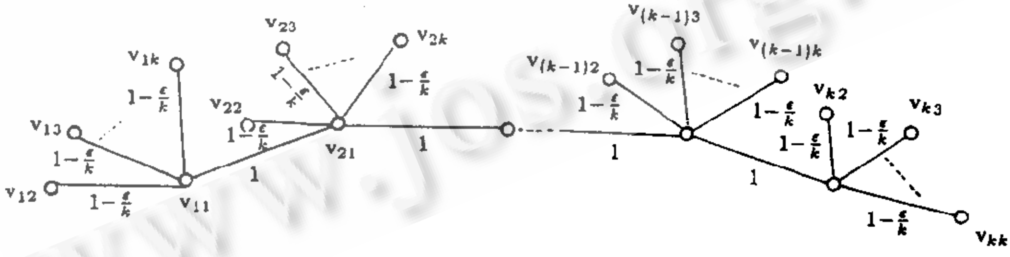
边　集 $E(k,\varepsilon)=\{(v_{p1},v_{pu})\mid p=1,2,\cdots,k$ 和 $q=2,3,\cdots,k\}\cup\{(v_{p1},v_{(p+1)1})\mid p=1,2,\cdots,k-1\}$.

其各边耗费规定如下：

$$c_{v_{p1},v_{pu}}=1-\frac{\varepsilon}{k}\quad,\quad 当\ p=1,2,\cdots,k\ 和\ q=1,2,\cdots,k.$$

$$c_{v_{p1},v_{(p+1)1}}=1\quad,\quad 当\ p=1,2,\cdots,k-1.$$

我们给出图 $G(k,\varepsilon)=G(V(k,\varepsilon),E(k,\varepsilon))$ 的图形如下：



由我们的 OGP 算法产生的划分为 $H=\{V_1,V_2,\cdots,V_k\}$，其中 $m=k$，只有 $W(V_1)=k-1$，其余的 $W(V_j)=0,(j=2,3,\cdots,k)$. 这样，划分的所有内部边耗费和 $W(H)=W(V_1)+W(V_2)+\cdots+W(V_k)=(k-1)+0+\cdots+0=k-1$.

现在我们考虑这样一种划分 $Q=\{Q_1,Q_2,\cdots,Q_k\}$，其中

$Q_i=\{v_{iu}\mid u=1,2,\cdots,k\}$，　$i=1,2,\cdots,k$

$$W(Q)=W(Q_1)+W(Q_2)+\cdots+W(Q_k)=(k-1)\cdot(1-\frac{\varepsilon}{k})+(k-1)(1-\frac{\varepsilon}{k})+\cdots+(k-1)(1-\frac{\varepsilon}{k})=k(k-1)(1-\frac{\varepsilon}{k})$$

© 中国科学院软件研究所　http://www.jos.org.cn

从而　$W(Q)/W(H)=\dfrac{k(k-1)(1-\frac{\varepsilon}{k})}{k-1}=k(1-\frac{\varepsilon}{k})=k-\varepsilon.$

即　$W(Q)/W(H)=k-\varepsilon$，也就是 $W(Q)=(k-\varepsilon)\cdot W(H)$，证毕.

3.3 算法的有效性

关于算法的时间复杂性分析. 设图 $G=(V,E)$ 有 $|V|=n$ 个结点，在第 1 步时共有$\binom{n}{k}$个 k—子集作为候选结点子集，第 2 步时共有$\binom{n-k}{k}$个，如此下去，第 m 步不会超$\binom{k}{k}$个 k—子集. 对于每个结点子集，确定其内部边耗费和的时间不会超过 $O(k^2)$. 因此，依算法 OGP 所产生的结点划分 $H=\{V_1,V_2,\cdots,V_m\}$，其完成时间不会超过

$$ck^2\binom{n}{k}+ck^2\binom{n-k}{k}+\cdots+ck^2\binom{k}{k}\quad=ck^2\left(\binom{n}{k}+\binom{n-k}{k}+\cdots+\binom{k}{k}\right)\leqslant ck^2n^k$$

这里 c 是一个常数，由此可知算法 OGP 的时间复杂度为 $O(k^2n^k)$，当 k 是给定常数时，该算法是多项式时间的有效算法.

参考文献

[1]Cook, S., "The Complexity of Theorem—Proving Procedures", Conference Record of Third ACM Symposium on Theory of Computing, 1970, 151-158.

[2]Richard M. Karp, "Reducibility among Combinatorial Problems", Complexity of Computer Computation, Plemum Press, New York, 1972, 85-103.

[3]Lin, S., "Heuristic Programming as an Aid to Network Design", Network 5, 1975, 33—43.

[4]B. W. Kernighan and S. Lin, "Efficient Heuristic for Partitioning Graphs", The Bell System Technical Journal, Vol. 49, February 1970, 291-307.

[5]鄢勇等，"图的最优 k 划分"，计算机学报，第 3 期，1990.

© 中国科学院软件研究所 http://www.jos.org.cn