Approximation algorithms for the generalized graph partitioning problems with restrictions

Jiping Liu[a*] Cheng Zhao[b] Jian Liang Zhou[c]

[a]Departmentt of Mathematics & Computer Sciences, University of Lethbridge,
Lethbridge, Alberta, Canada, T1K 3M4.

[b]Department of Mathematics and Computer Sciences, Indiana State University,
Terre Haute, IN 47809 USA.

[c]School of Information Science and Technology, University of Science & Technology of China,
Hefei, 230027, P.R. of China.

Abstract

We propose the following graph partitioning problem: given an edge-weighted graph $G = (V, E, w)$ and disjoint subsets U_p of V where $|U_p| \leq k$ for $1 \leq p \leq r$, find a minimum weighted set of edges $E' \subseteq E$ such that its removal separates the graph G into k parts and each part contains at most one vertex from each U_p for $1 \leq p \leq r$. We present three approximation algorithms and show that the approximation algorithms produce good approximations.

Key words: approximation algorithms, graph partitioning, NP-complete.

AMS subject classifications. 05C, 68R

1 Introduction

Distributed memory architectures are becoming increasingly popular as a result of promised scalability at reduced costs and the availability of high performance microprocessors. This architecture requires that data associated with a given computation be partitioned and distributed to the local storage of each individual processor. How this is done will affect the program performance. However, programming of distributed memory multiprocessors is difficult and error-prone due to the lack of a single uniform global address space. The research by Li and Chen [8], Banerjee, Eigenmann, Nicolau and Padua [1], Gupta and Banerjee [3], Knoble, Lukas and Steele [7], and Ramanujam and Sadayappan [11] has concentrated on automating this process. In [8], Li and Chen have modeled this process to the so called (primary) index domain alignment problem, and proved that it is NP-complete for the class of graphs with alignment dimension 2. In [4], [5], and [9], the authors generalized the index domain problem. In this paper we further generalize it to a more general graph partitioning problem.

The generalized minimum k-multiway cut problem with restrictions:

Let $G = (V, E, w)$ be an edge-weighted graph of order n, where $w : E \to \{0, 1, 2, \ldots\}$. Let U_p, $1 \leq p \leq r$, be disjoint subsets of V with at most k specified vertices (terminals) and at least one contains k vertices. Find a minimum weighted set of edges $E' \subseteq E$ such that the removal of E' separates the graph G into k parts and each part contains at most one vertex from each U_p, i.e., we need to find a partition $S = \{V_1, V_2, \ldots, V_k\}$ of V into k parts such that V_i contains at most one vertex of U_p and $w(S^C) = \sum_{i=1}^{k} \sum_{e \in E \setminus G[V_i]} w(e)$ is the minimum, where $G[V_i]$ is the induced graph on vertex set V_i for

*This research was partially supported by the NSERC of Canada.

$1 \le i \le k$.

When $V(G) = \cup_{p=1}^{r} U_p$ and each U_p is an independent vertex set for $1 \le p \le r$, this problem is reduced to the index domain alignment problem. In this case, the problem is shown to be NP-complete and the related local minimum is also hard to find (i.e., it is PLS-complete, see [4] for details).

When $r = 1$, this problem is the k-Multiway Cut problem [2]. This problem is MAX SNP-hard even for fixed $k \ge 3$ (see [2]), and it has applications in parallel and distributed computing as well as in chip design [12]. In [2], Dahlhaus, Papadimitriou, Seymour, and Yannakakis gave a simple algorithm guaranteed to produce solutions within a factor of $2(1-\frac{1}{k})$. Later, in [6], the performance ratio was improved to $1.3438 - \epsilon_k$.

Analogously, we have the following generalized maximum k-multiway cut problem with restrictions: given an edge-weighted graph $G = (V, E, w)$ with disjoint subsets U_p, $1 \le p \le r$ of V of sizes at most k, find a vertex partition $S = \{V_1, V_2, \ldots, V_k\}$ of V such that each V_i contains at most one vertex from each U_p, and $\sum_{i=1}^{k} \sum_{e \in E \setminus G[V_i]} w(e)$ is the maximum. This problem generalizes the Max k-Cut problem which has been proved to be MAX SNP-complete (see [10]).

2 Three approximation algorithms

Let $a = \sum_{e \in G[U_p], 1 \le p \le r} w(e)$. The **Local Search Algorithm** and the **Derandomized Algorithm** and the analysis presented in [9] can be modified to apply to the above generalized graph partition problem.

Theorem 1 *For the generalized minimum (maximum) k-multiway cut problem with restrictions, both the Local Search Algorithm and the Derandomized Algorithm produce feasible solutions S with weight at most (at least) $w(S^C) \le (1 - \frac{1}{k})w(G) + \frac{1}{k}a$.*

In the following, we present another approximate algorithm.

A Mixed Algorithm

Step 1: *run any well-known k partitioning algorithm (for example, the merge algorithm proposed in [5]) on the part $\cup_{p=1}^{r} U_p$ (view each U_p as an independent set). Let $\{V_1, V_2, \cdots, V_k\}$ be the partition.*

Step 2: *construct graph G' from G by shrinking each set V_i into a vertex v_i, replacing multiple edges by a single edge with weight the sum of the weights of the multiple edges. Then apply any k-Multiway Cut Algorithm to G' with k specified vertices $\{v_1, v_2, \ldots, v_k\}$.*

Modifying the analysis in [5], we can show that a feasible solution produced by the Mixed Algorithm is as good as those produced by the Local Search and Derandomized Algorithms.

3 A probabilistic model

In this section, we consider a special case: $V(G) = \cup_{p=1}^{r} U_p$. We further note that for any feasible partition S, all edges in $\cup_{p=1}^{r} E(U_p)$ are crossing edges and hence their weight are always counted in. Therefore, we may assume that $\cup_{p=1}^{r} E(U_p) = \emptyset$ and $a = \sum_{e \in G[U_p], 1 \le p \le r} w(e) = 0$. Thus, G is an r-partite graph. Moreover, by assigning 0-weight probability to some edges, we may assume that G is a complete r-partite graph.

It is natural now if we consider the following probabilistic model: Given an edge weighted complete r-partite graph $T_{kr,r}$ with k vertices in each of $U_1, U_2, \cdots, U_r$, let edge weights be chosen independently from $\{0, 1, 2, \cdots, \Delta\}$ and $Pr\{X = i\} = p_i$, for $i = 0, \cdots, \Delta$, where X is the weight between any two adjacent vertices and $\sum_{i=0}^{\Delta} p_i = 1$.

For any feasible partition $S = \{V_1, V_2, \ldots V_k\}$, let $w(S) = \sum_{e \in G[V_i], 1 \le i \le k} w(e)$. We note that each induced graph on V_i is a K_r.

Theorem 2 *Let $T_{kr,k}$ be an edge weighted complete random graph with partition $U_1, \cdots, U_r$. Let S be any feasible partition of the generalized minimum k-multiway cut problem with restrictions. Then*

$$\frac{w(T_{kr,r})}{k} \le w(S) \le \frac{w(T_{kr,r})}{k} + O(k^{\frac{3}{4}} r^{\frac{3}{2}})$$

with probability $1 - o(1)$ when $kr \to \infty$.

Proof: Recall that the partition S induces k disjoint K_r's, say $K_r^1, \cdots, K_r^k$. Let $\xi (= w(S))$ be the total weight of edges in these k cliques K_r's.

It is obvious that

$$\xi = \sum_{s=1}^{k} \Big(\sum_{i,j \in V(K_r^s)} X_{ij} \Big)$$

where X_{ij} represents the weight between adjacent vertices i and j. Since

$$P(\{X_{ij}=w\})=p_w,$$

for $w=0,\cdots,\Delta$, it follows that the expectation of ξ is

$$E(\xi)=\sum_{s=1}^{k}\Big(\sum_{i,j\in V(K_r^s)}E(X_{ij})\Big)=k\frac{r(r-1)}{2}E(X),$$

where $E(X)=\sum_{w=1}^{\Delta}wp_w$.

Also, the variance of ξ is

$$V(\xi)=\sum_{s=1}^{k}\Big(\sum_{i,j\in V(K_r^s)}V(X_{ij})\Big)=k\frac{r(r-1)}{2}V(X),$$

where $V(X)=\sum_{w=1}^{\Delta}w^2p_w-(E(X))^2$.

By Chebyshev's inequality, we know that

$$P\Big(|\xi-E(\xi)|\geq\sqrt{k^{1/2}rV(c)}\Big)\leq\frac{V(\xi)}{\Big(\sqrt{k^{1/2}rV(\xi)}\Big)^2}.$$

Notice that $\frac{V(\xi)}{\left(\sqrt{k^{1/2}rV(\xi)}\right)^2}=\frac{1}{k^{1/2}r}$. Therefore,

$$E(\xi)-\sqrt{k^{1/2}rV(\xi)}\leq\xi\leq E(\xi)+\sqrt{k^{1/2}rV(\xi)},$$

with probability $1-o(1)$ as kr approaches infinity. Since the expectation of $w(T_{kr,r})$ is

$$E(w(T_{kr,r}))=k\frac{kr(r-1)}{2}E(X)=kE(\xi),$$

and the variance of $w(T_{kr,r})$ is

$$V(w(T_{kr,r}))=k\frac{kr(r-1)}{2}V(X)=kV(\xi),$$

it follows that

$$\xi\geq\frac{1}{k}E(w(T_{kr,r}))-\sqrt{k^{-1/2}rV(w(T_{kr,r}))}\qquad(1)$$

$$\xi\leq\frac{1}{k}E(w(T_{kr,r}))+\sqrt{k^{-1/2}rV(w(T_{kr,r}))}\qquad(2)$$

with probability $1-o(1)$ for any arbitrary feasible solution as kr is big enough.

Applying Chebyshev's inequality again for $w(T_{kr,r})$, we obtain that

$$P\Big(|w(T_{kr,r})-E(w(T_{kr,r}))|\geq\sqrt{k^{\frac{1}{2}}rV(w(T_{kr,r}))}\Big)$$
$$\leq\frac{V(w(T_{kr,r}))}{\Big(\sqrt{k^{1/2}rV(w(T_{kr,r}))}\Big)^2}.$$

Therefore,

$$w(T_{kr,r})\geq E(w(T_{kr,r}))-\sqrt{k^{1/2}rV(T_{kr,r})}\qquad(3)$$

and

$$w(T_{kr,r})\leq E(w(T_{kr,r}))+\sqrt{k^{1/2}rV(T_{kr,r})},\qquad(4)$$

with probability $1-o(1)$ as kr is big enough. From (1)-(4), we have

$$\begin{aligned}w(S)&\leq\frac{E(w(T_{kr,r}))-\sqrt{k^{1/2}rV(T_{kr,r})}}{k}\\&\quad+\frac{\sqrt{k^{1/2}rV(w(T_{kr,r}))}}{k}\\&\quad+\sqrt{k^{-1/2}rV(w(T_{kr,r}))}\\&\leq\frac{w(T_{kr,r})}{k}+\frac{\sqrt{k^{1/2}rV(w(T_{kr,r}))}}{k}\\&\quad+\sqrt{k^{-1/2}rV(w(T_{kr,r}))}\\&\leq\frac{w(T_{kr,r})}{k}+O(k^{3/4}r^{3/2})\end{aligned}$$

with probability $1-o(1)$. This completes the proof of Theorem 2.

Corollary 1 *Let $G=T_{kr,r}$ be an edge weighted complete r-partite random graph defined above. Let S be an optimum partition of the generalized minimum k-multiway cut problem with restrictions. Then with higher probability,*

$$\lim_{kr\to\infty}\frac{kw(S)}{w(T_{kr,r})}=1.$$

Proof: Note that $w(T_{kr,r})$ is equal to $k^2r^2E(X)$, the corollary follows from Theorem 2.

We have similar results for the generalized maximum k-multiway cut problem with restrictions.

Theorem 3 *Let $T_{kr,k}$ be an edge weighted complete random graph with partition $U_1, \cdots, U_r$. Let S be any optimum partition of the generalized maximum k-multiway cut problem with restrictions. Then*

$$\frac{w(T_{kr,r})}{k} - O(k^{\frac{3}{4}} r^{\frac{3}{2}}) \leq w(S) \leq \frac{w(T_{kr,r})}{k}$$

with probability $1 - o(1)$ when $kr \to \infty$, and thus

$$\lim_{kr \to \infty} \frac{kw(S)}{w(T_{kr,r})} = 1.$$

Remark: Corollary 1 and Theorem 3 suggest that if a weighted graph G with specified sets $U_1, U_2, \ldots, U_r$ and $V(G) = \cup_{p=1}^{r} U_p$ is an instance of the generalized minimum (maximum) k-multiway cut problem with restrictions, then the optimum solution of the problem is asymptotically $(1 - \frac{1}{k})w(G) + \frac{1}{k}a$ if $|V(G)|$ is large.

4 Conclusion

In this paper, we proposed a most general graph partition problem with restrictions which generalizes some well-known graph partition problems. We also proposed some approximation algorithms for the problem. By introducing a probabilistic model, we showed that solutions produced by the local search, the derandomized, and the mixed algorithms are good approximations.

References

[1] U. Banerjee, R. Eigenmann, A. Nicolau and D. A. Padua, Automatic Program Parallelization, Proceedings of the IEEE, Vol. 81, No. 2, 1993, pp. 211-243.

[2] E. Dahlhaus, D. S. Johnson, C. H. Papadimitriou, P. D. Seymour, and M. Yannakakis, The complexity of multi-terminal cuts, SIAM, J. Comput., 23(1994), 864-894.

[3] M. Gupta and P. Banerjee, Demonstration of automatic data partitioning techniques for parallelizing compilers on multicomputers, IEEE Transactions on Parallel and Distributed Systems, No. 2, 3 (1992), pp. 179-193.

[4] G. He, J. Liu and C. Zhao, Some graph partitioning problems, J. Combin. Information and System Science, vol 24 (1999), no. 2-4, pp. 177-194.

[5] G. He, J. Liu and C. Zhao, Approximation algorithms for some graph partitioning problems, Journal of Graph Algorithms and Applications, vol. 4 no. 2 (2000) 1 – 11.

[6] D. R. Karger, P. Klein, C. Stein, M. Thorup, and N. E. Young, Rounding algorithms for a geometric embedding of minimum multiway cut. Math. Oper. Res. 29 (2004), no. 3, 436–461.

[7] K. Knoble, J. Lukas and G. L. Steele, Data optimization, allocation of arrays to reduce communication in SIMD Machines, Journal of Parallel and Distributed Computing, 8 (1990), pp., 102-118.

[8] J. Li and M. Chen, Index domain alignment: Minimizing cost of cross-reference between distributed arrays, in Proceedings of the third Symposium on Frontiers of Massively Parallel Computation, College Park, October, 1990, pp. 424-433.

[9] J. Liu, Y. J. Peng, and C. Zhao, Generalized k-Multiway Cut Problems, submitted.

[10] C. H. Papadimitriou and M. Yannakakis, Optimization, approximation and complexity classes, J. of Computer and System Sciences, 43 (1991), pp 425-440.

[11] J. Ramanujam and P. Sadayappan, Compile-Time Techniques for Data Distribution Memory Machines, IEEE Transactions on Parallel and Distributed Systems, No.4, 2 (1991), pp. 472-483.

[12] H. S. Stone, Multiprocessor scheduling with the aid of network flow algorithms, IEEE Trans. Software Eng. 3 (1977), 85-93.