Approximations via Partitioning

Magnús M. Halldórsson

14 March 1995

Abstract

We consider the approximation of weighted maximum subgraph problems by partitioning the input graph into easier subproblems. In particular, we obtain efficient approximations of the weighted independent set problem with performance ratios of $O(n(\log\log n/\log n)^2)$ and $(\Delta+2)/3$, with the latter improving on a $\Delta/2$ ratio of Hochbaum for $\Delta \geq 5$. We also obtain a $O(n/\log n)$ performance ratio for various maximization problems where a subset of a solution is also a solution.

1 Partitioning and hereditary induced subgraph problems

A property of graphs is *hereditary* if whenever it holds for a graph it also holds for its induced subgraphs. For a hereditary property, the associated subgraph problem is that of finding a subgraph of maximum weight satisfying the property. We say that a problem is *approximable within* $f(n)$ if there is a polynomial time algorithm that on graphs with n vertices returns a feasible solution within $f(n)$ factor of optimal.

Hereditary can be generalized to other discrete structures. A property is hereditary if whenever it holds for a subset X of the instance, it also holds for any subset of X.

The main purpose of this note is to study the implications of the following lemma. We state it in the language of graphs, while it can also be applied to other hereditary problems.

Proposition 1 *Let Π be the problem of finding a maximum induced subgraph satisfying a hereditary property π. If we can partition the graph G into subgraphs $G_1, G_2, \ldots G_t$ and solve Π optimally on each G_i, then we can approximate Π on G within t.*

Our results We present several applications of this approach in Section 2. The first is to partition the input into small bits, each of which can be searched exhaustively for an optimal solution. This suffices to obtain a performance ratio of $O(n/\log n)$ for various problems for which a subset of any solution is also a solution, including hereditary induced subgraph problems, and the problems *Longest Common Subsequence* and *Maximum Satisfying Linear Subsystem* (see [8] for statements of these problems and references). Here, n is the total number of items in the instance, as given in the measure. This also holds for weighted versions of these problems. We strengthen this approximation slightly for unweighted versions to show how to find a solution of size $\log_{2n/OPT} n$, where OPT is the size of the optimal solution.

We then obtain a stronger performance ratio of $O(n(\log\log n/\log n)^2)$ for the *Weighted Independent Set* problem, and other hereditary problems with a forbidden clique or independent set. This is obtained by partitioning the graph into subgraphs that are either independent sets or unions of at most $\log n/\log\log n$ cliques.

We then consider graphs of maximum degree Δ. We present an efficient algorithmic proof of an obscure lemma of Lovász, and use it to approximate weighted hereditary induced subgraph problems within $\lceil(\Delta+1)/3\rceil$. An additional technique allows us to improve the bound for the *Weighted Independent Set* problem to $(\Delta+2)/3$, for $\Delta \geq 5$, from the $\Delta/2$ ratio of Hochbaum [11].

Finally, we consider some limitations on this technique of approximating by partitioning the input into easy subproblem. In particular, we find that partitioning into subgraphs with any given hereditary property cannot yield a performance ratio better than what we obtained.

A more general theorem Proposition 1 is a special instance of the following more general theorem, which we shall also refer to.

Theorem 2 *Let Π be a hereditary, weighted induced subgraph problem. Suppose we can:*

1. *extract t induced subgraphs of G, $G_1, G_2, \ldots G_t$, such that each node in G is contained in at least k different G_i, and*
2. *find feasible Π-solutions $HEU(G_i)$ such that $HEU(G_i)\rho_i \geq OPT(G_i)$.*

Then,

$$HEU(G) = \max_i HEU(G_i) \geq OPT(G) \cdot k/(\sum_i \rho_i).$$

Proof. Restricting the optimal solution on G to each subgraph yields a feasible solution, thus

$$\sum_i OPT(G_i) \geq k \cdot OPT(G).$$

Some G_j must have the property that

$$OPT(G_j) \geq \frac{\rho_j}{\sum_i \rho_i} k \cdot OPT(G).$$

Then,

$$HEU(G_j) \geq OPT(G_j)/\rho_j \geq OPT(G)/(\sum_i \rho_i).$$

∎

Previous approximation results via partitioning Two notable applications of Theorem 2 in the literature deal with geometric graphs. Baker [2] approximated a host of problems on planar graphs by partitioning into K-outerplanar graphs and developing dynamic programming methods for solving the problems on these graphs. Hunt et al. [12] similarly approximated independent sets in unit-disk graphs by partitioning into graphs of bounded treewidth.

2 Applications

2.1 Partition into small subsets

We say that a property of graphs is *EXP-checkable* if, given a graph on n vertices, the property can be verified in time at most 2^{n^c}, for some constant c. The class of hereditary EXP-checkable properties includes all the common ones, such as the ones listed in [17]. However, it probably doesn't include such esoteric properties as:

> Subgraph of a minimum cardinality graph containing no s-clique and no t-independent set.

Suppose we arbitrarily partition a graph into $n/\log n$ sets each with $\log n$ vertices. Then, simple brute-force suffices to find an optimal solution to any hereditary maximum subgraph problem in polynomial time.

Proposition 3 *Any EXP-checkable hereditary, weighted induced subgraph problem can be approximated within* $n/\log n$*.*

Surprisingly, this $n/\log n$-approximation appears to be the best that is known for most such problems. A property is *non-trivial* if it holds for some and fails for some graphs. It is known that, the subgraph problem for any non-trivial hereditary property cannot be approximated within any constant unless $P = NP$, and stronger results are known for most properties [17].

This can be generalized to some hereditary subset problems.

Proposition 4 *Let* Π *be a hereditary subset problem for which the question of feasibility question of an instance with measure at most* $\log n$ *can be answered in polynomial time. Here,* n *is an upper bound on the measure of the instance. Then, the problem can be approximated within* $O(n/\log n)$*.*

We previously applied this approach to the problems *Maximum Common Pointset* and *Maximum Common Subtree* [1]. Refer to [8] for a definitions and references on these and the following problems. It also applies to *Longest Common Subsequence*, which we can approximate within $n \log n$, where n is the length of the shortest sequence.

Another problem is *Maximum Satisfying Linear Subsystem*, defined as follows: Given a system $Ax = b$ of linear equations, with A an integer $m \times n$ matrix and b an integer m vector, find a rational vector $x \in \mathcal{Q}^n$ that satisfies the maximum number of equations. Since feasibility of a given system can be solved in polynomial time via linear programming, we can approximate this problem via Proposition 1 within $O(m/\log m)$ (or $m \log N$, where N is the size of the input). This holds equally if the variables are restricted to take on a binary values or if equality is replaced by inequalities ($>$, $\geq$). It also holds if some additional constraints/equations are required to be satisfied by a solution.

2.2 Partition into searchable subsets

For unweighted problems we can obtain a better approximation, although the performance ratio is not directly affected. The following algorithm was originally used by Berger and Rompel [4] as a part of an approximate graph coloring algorithm.

Theorem 5 *Let* Π *be a hereditary, EXP-checkable, induced subgraph problem. Given a graph* G*, let* OPT *be the size of an optimal solution, and let* k *denote* $2n/OPT$*. Then, we can obtain a* π*-subgraph of size* $\log_k n$ *in polynomial time.*

Proof. Suppose we are given the value of k. We can obtain this value by applying binary search on k until the following procedure yields a proper solution. Let $m = k \log_k n$.

Partition the graph into n/m induced subgraphs, containing m vertices each (ignoring ceilings). At least one of these subgraphs contains a $\log_k n$ size subset of the optimal solution, and we can find such a solution by searching through all the subsets of size at most $\log_k n$ in each partition.

The number of such subsets is

$$n/m \cdot \binom{m}{\log_k n} \;\leq\; n(ek)^{\log_k n}/m \;\leq\; n^3/m,$$

hence the algorithm runs in $O(n^3)$ time. ∎

This approach also applies to the hereditary subset problems discussed earlier. Also, observe that this approach above can produce a $\log_k n$ size subset that satisfies all checkable properties of a subset of an optimal solution. This property proved useful for improving the performance ratio known for graph coloring [10].

2.3 Partition into independent sets and union of cliques

A theorem of Erdős and Szekeres [9] on Ramsey numbers yields an efficient algorithm for finding either cliques or independent sets of non-trivial size.

Theorem 6 (Erdős, Szekeres) *Any graph on n vertices contains a clique on t vertices and independent set on s vertices such that $\binom{s+t-2}{s-1} \geq n$.*

Theorem 6 implies that we can obtain either an independent set of size $\log^2 n$, or a clique of size $\log n / \log\log n$. We now form a partition where each class is either an independent set, or a (not necessarily disjoint) union of $\log n / \log\log n$ different cliques. This yields a partition into $O(n(\log\log n/\log n)^2)$ classes.

The weighted independent set problem can be solved on the latter type by exhaustively checking all $(\log n/\log\log n)^{\log n/\log\log n} = O(n)$ possible combinations of selecting one vertex from each clique.

Corollary 7 *The weighted independent set problem can be approximated within $O(n(\log\log n/\log n)^2)$.*

Similar results, within constant factors, hold for induced subgraph problems whose properties are not satisfied either by some independent set or some clique. For comparison, the best performance ratio known for the unweighted case is only slightly better or $O(n/\log^2 n)$ [5]. It is known that the unweighted problem cannot be approximated within n^t, where $t \geq 1/6 - \delta$, unless $P = NP$ [3], and it has been conjectured that $t = o(1)$ is still impossible.

2.4 Partition into low-degree subgraphs

A little-known lemma of Lovász provides the tool for partitioning a graph into subgraphs of low maximum degree.

Lemma 8 (Lovász [15]) *Let $G = (V, E)$ be a multigraph with no self loops. Let $k_1, k_2, \ldots, k_t$ be non-negative integers such that $1 + (\sum_i k_i - 1) = \Delta(G)$. Then, V can be partitioned into t subsets inducing subgraphs $G_1, G_2, \ldots G_t$ such that $\Delta(G_i) \leq k_i$, for $i = 1, 2, \ldots t$.*

Proof. Lovász's proof of this lemma involves a common local search strategy that repeatedly applies the following rule: If some vertex in G_i has more than k_i neighbors, then move it to some G_j where the vertex has at most k_j neighbors. By the pigeon-hole principle, at least one of the subgraphs must have this property, and his proof shows that this strategy terminates.

We show that at most $3m + n$ iterations suffice. We consider each iteration to consist of two operations. First, a vertex v is moved out of its current subgraph G_i, thereby decreasing the number of edges within that subgraph by at least $k_i + 1$. Then, it is moved into another subgraph G_j, increasing the number of edges within that class by at most k_j. Consider the set of operations performed in the first (up to) $3m + n$ iterations, and pair together move-in and move-out operations involving the same class. The effect of each pair of operations is a net drop in the number of edges within that subgraph. There are at most n insertion operations that cannot be paired, consisting of distinct vertices $v_1, v_2, \ldots, v_t$, $t \leq n$. The number of edges added to the subgraph due to these operations is at most $\sum_{i=1}^{t} d(v_i) \leq 2m$. Also, there are at most m edges initially within the subgraphs. It follows that if these movements continue until the $3m+n$-th iteration, no edges can remain within the subgraph after that iteration, and the process terminates.

The algorithm can be implemented by pre-computing the degrees of each vertex into each class of the starting partitition, with each move requiring at most $O(\Delta)$ updates to this data. The complexity is therefore bounded by $O(\Delta(m+n))$. ∎

A corollary of this lemma is that we can partition a graph into at most $\lceil(\Delta+1)/3\rceil$ graphs of degree at most 2. Any hereditary induced subgraph problem can be solved efficiently on such graphs via dynamic programming. Any property π holds either for every path, or for all paths of length up to t, where t is a fixed constant, and the same dichotomy holds for cycles.

Theorem 9 *Hereditary weighted induced subgraph problems can be approximated within* $\lceil(\Delta+1)/3\rceil$.

The previous best approximation for the weighted independent set problem is $\Delta/2$ due to Hochbaum [11]. We can also use here approximation for $\Delta = 3$ to improve our ratio in the case $\Delta \pmod 0 \equiv 0$. In this case, we partition into $\Delta/3$ classes, where all but the last have maximum degree 2, and the last class of maximum degree 3 is approximated within 3/2.

Corollary 10 *The weighted independent set can be approximated within* $(\Delta+2)/3$*, and within* $(\Delta+1)/3$ *when* $\Delta \equiv 2 \pmod 3$.

By applying a preprocessing method of Hochbaum [11], we can obtain improved approximations of weighted vertex cover, for $\Delta \geq 5$.

Corollary 11 *The weighted vertex cover problem can be approximated within* $2 - 3/(2\Delta+4)$*, and within* $2 - 3/(2\Delta+2)$ *when* $\Delta \equiv 2 \pmod 3$.

Lovász's lemma also has implications for the coloring of bounded-degree graphs, observed independently by Catlin [7], Borodin and Kostochka [6] and Lawrence [14], and recently rediscovered by Lau [13] and communicated to this author.

Partition the graph into subgraphs of degree 3 or 4, with $(\Delta+2) \pmod 3$ subgraphs of degree 4 and the remainder of degree 3. Assuming the graph contains no complete graph on 4 vertices, each subgraph can be colored with $\Delta(G_i)$ colors by the algorithm that follows from the constructive proof of Brooks' theorem due to Lovász [16].

Proposition 12 *Graphs without 4-cliques can be colored with* $(3\Delta+2)/4$ *colors in time* $O(\Delta(m+n))$.

2.4.1 Partition into bipartite subgraphs

Hochbaum [11] proved the following theorem.

Theorem 13 (Hochbaum) *If we can color a graph G with k colors, we can approximate the weighted independent set within $k/2$.*

We can obtain the same result by pairing the color classes together into classes of bipartite graphs. When k is odd, the last three are grouped together and each of the three pairs is solved optimally, with the maximum attaining a ratio of $3/2$; the combined approximation ratio is then $(k-3)/2 + 3/2 = k/2$.

Although based on different principles, the approach of [11] is also based on computing the maximum independent set in a bipartite graph. The advantage with our approach is that when the classes are of roughly equal size, the sum of the complexity of our method will be smaller than the complexity of a single execution of the whole. In the weighted case, we may save up to a factor of k over the $O(mn \log n)$ complexity of [11] if the color classes are of roughly equal sizes.

In the unweighted case, we can guarantee time savings of a factor of $\sqrt{k}$ over the $O(m\sqrt{n})$ complexity of [11] by simply using 2-coloring on classes with more than n/k vertices. Another advantage of this approach is that unlike the approach of [11], it yields the same approximation for other induced subgraph problems that are solvable on bipartite graphs. We note, however, that the above method alone does not yield the improvement for the vertex cover problem that the approach of [11] does.

2.5 Limitations of partitioning

The wide applicability of this partitioning technique might offer a glimmer of hope for approximating the independent set problem in general graphs within $n^{1-\epsilon}$, for some $\epsilon > 0$. The following observation casts a shade on that proposal.

For a property Π, the Π-chromatic number of a graph is the minimum number of classes that the vertex set can be partitioned into such that the graph induced by each class satisfies Π. Scheinerman [18] has shown that for any non-trivial hereditary property Π, the Π-chromatic number of a random graph approaches $\theta(n/\log n)$. This indicates that our results are essentially the best possible.

Acknowledgement

I thank H. C. Lau for rediscovering a case of Lovász's lemma, telling me about it, and thereby leading me to consider partitions of graphs.

References

[1] T. Akutsu and M. M. Halldórsson. On the approximation of largest common point sets and largest common subtrees. In *Proc. Fifth Intl. Symp. on Algorithms and Computation*, pages 405–413, Beijing, China, Aug. 1994. Springer-Verlag LNCS #834.

[2] B. S. Baker. Approximation algorithms for NP-complete problems on planar graphs. *J. ACM*, 41:153–180, Jan. 1994.

[3] M. Bellare and M. Sudan. Improved non-approximability results. In *Proc. 26th Ann. ACM Symp. on Theory of Computing*, pages 184–193, May 1994.

[4] B. Berger and J. Rompel. A better performance guarantee for approximate graph coloring. *Algorithmica*, 5(4):459–466, 1990.

[5] R. B. Boppana and M. M. Halldórsson. Approximating maximum independent sets by excluding subgraphs. *BIT*, 32(2):180–196, June 1992.

[6] O. V. Borodin and A. V. Kostochka. On an upper bound of a graph's chromatic number, depending on the graph's degree and density. *J. Combin. Theory Ser. B*, 23:247–250, 1977.

[7] P. A. Catlin. A bound on the chromatic number of a graph. *Discrete Math.*, 22:81–83, 1978.

[8] P. Crescenzi and V. Kann. A compendium of NP optimization problems. Dynamic on-line survey available at `nada.kth.se`, Dec. 1994.

[9] P. Erdős and G. Szekeres. A combinatorial problem in geometry. *Compositio Math.*, 2:463–470, 1935.

[10] M. M. Halldórsson. A still better performance guarantee for approximate graph coloring. *Inf. Process. Lett.*, 45:19–23, 25 January 1993.

[11] D. S. Hochbaum. Efficient bounds for the stable set, vertex cover, and set packing problems. *Disc. Applied Math.*, 6:243–254, 1983.

[12] H. B. Hunt III, M. V. Marathe, V. Radhakrishnan, S. S. Ravi, D. J. Rosenkrantz, and R. E. Stearns. A unified approach to approximation schemes for NP- and PSPACE-hard problems for geometric graphs. In *Proc. of the Second European Symposium on Algorithms*, 1994.

[13] H. C. Lau. Local search for CSP and its applications to rescheduling and graph coloring. Technical Report No. 94TR-0014, Computer Science Department, Tokyo Inst. of Technology, 1994.

[14] J. Lawrence. Covering the vertex set of a graph with subgraphs of smaller degree. *Discrete Math.*, 21:61–68, 1978.

[15] L. Lovász. On decomposition of graphs. *Stud. Sci. Math. Hung.*, 1:237–238, 1966.

[16] L. Lovász. Three short proofs in graph theory. *J. Combin. Theory Ser. B*, 19:269–271, 1975.

[17] C. Lund and M. Yannakakis. The approximation of maximum subgraph problems. In *ICALP '93*, July 1993.

[18] E. R. Scheinerman. Generalized chromatic numbers of random graphs. *SIAM J. Disc. Math.*, 5(1):74–80, Feb. 1992.