An Approximation Algorithm for Complete Partitions of Regular Graphs

Magnús M. Halldórsson*

18 February 2004

Abstract

A complete partition of a graph is a partition of the vertex set such that any two classes are connected by an edge. We consider the problem of finding a complete partition maximizing the number of classes. This relates to clustering into the greatest number of groups so as to minimize the diameter (inter-cluster connectivity) without concern for the intra-cluster topology.

We give a randomized algorithm that approximates the complete partitioning number within a factor of $O(\sqrt{\log n})$ on regular graphs.

1 Introduction

A *complete partition* of a graph $G = (V, E)$ is a partition of the vertex set such that there is an edge between members of any two classes. That is, a partition $V_1, \ldots, V_t$ of V is complete if, for each i, j, $i \neq j$, there is an edge (v_i, v_j) such that $v_i \in V_i$ and $v_j \in V_j$. The CP problem is that of finding a complete partition with maximum number of classes. Let $cp(G)$ denote the maximum number of classes in a complete partition of a graph G.

One easy observation is that since there must be an edge between any pair of classes,

$$cp(G) \leq q(m),$$

where $m = |E(G)|$ and $q(m) \approx \sqrt{2m}$ is the largest value t such that $\binom{t}{2} \leq m$.

Motivation: Our complete partition problem falls under the general family of clustering problems with inter-cluster constraints (for example, each cluster should be an independent set) and additionally, some constraints on the relation between different clusters.

An important example is that of *low diameter decomposition*. There, the constraint is that the diameter of every cluster is "low" and that the number of edges between clusters is "low". Low diameter partitions have numerous applications in, e.g., synchronization in distributed computation, on-line tracking of mobile users, etc (see [AP90]).

The interplay between intra- and inter-cluster topology yields widely different considerations. Separating these concerns may provide basic understanding that can be translated back to the combined problems. In many ways, problems involving intra-cluster topology alone have been well studied in the literature. Namely, we obtain graph partitioning problems where each class should

*Department of Computer Science, University of Iceland, IS-107 Reykjavík, Iceland. **Email:** `mmh@hi.is`

satisfy some property Π. The concern of the current paper is to study a basic problem involving only the inter-cluster topology.

The motivation for the particular constraints as suggested in the complete partition problem comes from the goal of allowing fast communication in presence of large number of processors in a distributed setting.

Assume that the given graph represent the current communication lines between existing processors. Breaking the graph into "many" clusters implies that typically a cluster is of small size. This makes possible to cheaply add some fast means of communication (such as a bus) between vertices of the same cluster. In addition, if the partition is complete, fast communication between different clusters is possible using already existing communication lines. This can be useful in speeding the execution of important basic primitives such as broadcast.

Our results: We give the first bounds on the approximability of the complete partitioning problem. We present (in Section 2) a randomized algorithm that finds a complete partition of size $\Omega(q(m)/\sqrt{\log n})$ on regular graphs. Since $q(m)$ is an upper bound on the complete partitioning number $cp(G)$, this yields an absolute approximation factor of $O(\sqrt{\log n})$. We also show (in Section 4) that for random graphs with m edges, $cp(G) = O(q(m)/\sqrt{\log n})$, implying that the gap of a $\theta(\sqrt{\log n})$-factor is tight.

Related work: The cardinality of a maximum complete partition has been studied in graph theory as the *pseudo-achromatic number*. It was introduced by Gupta [G69], and shown NP-hard by Knisely [K92]. Before discussing these results further, we first introduce some related problems, citing only the most relevant results.

A coloring is a partition into independent sets, and a complete coloring is a complete partition into independent sets. The minimum (maximum) number of colors in a complete coloring of a graph is its *chromatic number* (*achromatic number*). The achromatic number is a lower bound on the pseudo-achromatic number, and is particularly relevant to this work. The *harmonious* number is the minimum number of colors in a coloring such that there is at most one edge connecting any pair of color classes. A coloring that is both complete and harmonious uses at most $q(m)$ colors, and thus forms also an optimal complete partition.

Cairnie and Edwards [CE97] showed that the achromatic number is NP-hard for trees by showing that it is NP-hard to determine if a graph contains a complete harmonious coloring. It therefore also yields the NP-hardness for CP of trees. Bodlaender's [Bod89] proof of the NP-hardness for graphs that are simultaneously cographs and interval graphs has the same property and also yields hardness for CP. Cairnie and Edwards [CE98] gave an algorithm that finds a complete coloring of trees having bounded degree with at least $q(m) - 1$ colors. Thus, CP also has an approximation with an additive term of 1.

There are few maximum partitioning problems known that do not involve colorings. A *domatic partition* is a partition of the vertex set into dominating sets. Feige et al. [FHKS00] gave a randomized algorithm that finds a domatic partition into $\delta(1-o(1))/\log n$ classes, while any such partition contains at most δ classes (δ is the minimum degree of the graph). It was also shown that it was hard to approximate the problem within a factor of $(1-\epsilon)\log n$, for any $\epsilon > 0$.

2 Algorithms for regular graphs

This section is devoted to the following result.

Theorem 2.1 *There is a polynomial time randomized algorithm that finds a complete partition of a d-regular graph into $\Omega(\sqrt{dn/\log n})$ classes.*

The algorithm can be made deterministic using the method of conditional expectation. Note that for G regular, we have an easy upper bound of $cp(G) \le q(m) \le \sqrt{dn}$.

Corollary 2.2 *For G regular, $\Omega(q(m)\sqrt{\log n}) \le cp(G) \le q(m)$.*

Let $t = c\sqrt{dn/\ln n}$ for a constant c to be determined later. Randomly partition the vertex set $V(G)$ into $2t$ equal-sized classes: $S_1, S_2, \ldots, S_t$ and $T_1, T_2, \ldots, T_t$. Let $S = \cup_i S_i$ and $T = \cup_i T_i$.

We also introduce the following notation. Let $n_S = |S|$ and $n_T = |T|$, so $n = n_S + n_T$. Let d_T (d_S) be the degree of vertices in T (S), respectively. Write $t = c'\sqrt{d_T n_T/\ln n_T}$.

We say that S_i is *expanding* if the size of the neighborhood of S_i in T, $N_T(S_i) = N(S_i) \cap T$, is at least $(1 - e^{-d_T/4t})n_T$. Let $\mathcal{C} = \{S_i \cup T_i | S_i \text{ is expanding}\}$. Observe that $\mathcal{C}$ can be computed in polynomial time. We show that with some constant probability, $\mathcal{C}$ is a complete partition of at least $\Omega(t)$ classes.

Let $\tilde{t} = \mathcal{C}$ be the random variable denoting the number of expanding sets S_i.

Lemma 2.3 $\Pr\left[\tilde{t} \ge \frac{t}{4}\right] \le \frac{1}{2}$.

Thus with probability at least half, $\mathcal{C}$ contains at least $t/4$ sets. Before proving this lemma, and the following lemma, let us continue to prove our main result.

We write $S_i \sim T_j$ to denote that there is an edge of G connecting S_i and T_j. Let A_{ij} denote the event that $S_i \not\sim T_j$ conditioned on S_i being an A-expander.

Lemma 2.4 $\Pr[A_{ij}] \le \frac{1}{t^2}$.

Since there are at most $\binom{t}{2} < t^2/2$ pairs of sets in $\mathcal{C}$, Lemma 2.4 implies that with probability at least $1/2$, the sets in $\mathcal{C}$ are pairwise adjacent. We arbitrarily add to these sets any vertices not already included in them. Thus, using Lemma 2.3, the resulting $\mathcal{C}$ is a complete partition of size at least $t/5$, with probability at least $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. It remains to prove Lemmas 2.3 and 2.4.

Proof of Lemma 2.3: To prove the lemma, it suffices to show the following claim.

Claim 2.5 *The probability that a given S_i is expanding is at least $1/2$.*

The claim implies that

$$\mathrm{E}\left[\tilde{t}\right] \ge t/2.$$

which by Markov inequality yields the statement of the lemma.

It remains to prove the claim. Let S_i be one of the sets randomly chosen from S, and let w be a vertex in T. In order for w not to be contained in $N_T(S_i)$, all the d_T neighbors of w must fail to be chosen for S_i. That is,

$$\Pr[w \notin N_T(S_i)] = \binom{|\overline{S_i}|}{d_T} \Big/ \binom{n_S}{d_T} \le (1 - 1/t)^{d_T} \le e^{-d_T/t}.$$

We divide the rest of the proof into two cases.

Case $d \ge 2t$: Then, the expected size of $N_T(S_i)$ is at least

$$\mathrm{E}[|N_T(S_i)|] \ge (1 - e^{-d_T/t})n_T.$$

Since n_t is the maximum possible size of $N_T(S_i)$, we have by Markov that

$$\Pr\left[|N_T(S_i)| \geq (1 - 2e^{-d_T/t})n_T\right] \geq \frac{1}{2}.$$

Since $2e^{-d_T/t} \leq e^{-d_T/2t}$, this says that the probability that S_i is *expanding* is at least half, as claimed.

Case $d \leq 2t$: Now,

$$\Pr\left[w \notin N_T(S_i)\right] \leq e^{-d_T/t} \leq 1 - d_T/2t.$$

The expected size of $N_T(S_i)$ is then at least

$$\mathrm{E}\left[|N_T(S_i)|\right] \geq \frac{d_T n_T}{2t},$$

while the maximum size of $N_T(S_i)$ can be bounded by $d_S|S_i| = d_S n_S/t = d_T n_T/t$. By Markov again,

$$\Pr\left[|N_T(S_i)| \geq \frac{d_T n_T}{4t}\right] \geq \frac{1}{2}.$$

Since $d_T/t \geq 1 - e^{d_T/t}$, this implies that the expansion probability is at least half, as claimed. □

Proof of Lemma 2.4: *Case $d_T \geq t$:* For an expanding set S_i,

$$\Pr\left[A_{ij}\right] = \binom{|\overline{N_T(S_i)}|}{|T_j|} / \binom{n_T}{|T_j|} \leq \left(\frac{|\overline{N_T(S_i)}|}{n_T}\right)^{|T_j|}.$$

Recalling the size of T_j and the definition of expansion, we have that

$$\Pr\left[A_{ij}\right] \leq (e^{-d_T/4t})^{n_T/t} = e^{-\log n/4c^2} = n^{1/(4c^2)}.$$

Thus, the lemma holds for any $c \leq \sqrt{1/8}$. □

2.1 Upper bound on $cp(G)$:

We propose a new upper bound on the optimal solution size $cp(G)$ that applies also to non-regular graphs. The bound $q(m)$ is a good approximation of $cp(G)$, when G is regular, but can be far off when G is non-regular. For instance, when G is the complete bipartite graph $K_{d,n-d}$, $q(m) \approx \sqrt{dn}$, while $cp(G) = d + 1$. Thus, $q(m)$ can differ by a factor of as much as $\sqrt{n}$.

Definition 2.1 *A subgraph $H \subseteq G$ is* degree-stunted *if $\Delta(H) < q(|H|)$.*

Define $q^*(G)$ to be the largest value $q(|H|)$ of a degree-stunted subgraph H. The subgraph H with $q(|H|) = q^*(G)$ is the maximum degree-stunted subgraph of G.

To see that $q^*(G)$ forms an upper bound on $cp(G)$, consider a maximum complete partition $V_1, \ldots, V_{cp(G)}$. Select arbitrary $\binom{cp(G)}{2}$ edges $e_{i,j}$, $1 \leq i < j \leq cp(G)$ such that $e_{i,j}$ connects vertices in V_i and V_j, and let H be the subgraph of G formed by the edges $e_{i,j}$. Then, $q(|H|) = cp(G)$ while the degree of each vertex is at most $cp(G) - 1$, so H is degree-stunted.

Intuitively, at most $t-1$ edges incident on a vertex can ever contribute to a complete t-partition; thus, we may safely ignore the other edges.

The proof of the following claim is omitted. The computation involves a reduction to a maximum matching problem.

Claim 2.6 *A maximum degree-stunted subgraph of a graph G can be found in polynomial time.*

3 Random graphs

This section is devoted to the following result that shows that the $\sqrt{\log n}$ gap of our randomized algorithm is unavoidable.

Theorem 3.1 *Let $G = G_{n,p}$ be a random graph. Let $m = \binom{n}{2}p$. Then, $cp(G) = O(q(m)/\sqrt{\log n})$, with high probability.*

Proof: Let $t = C\sqrt{m/\ln n}$, for C large constant. Let $d = pn = 2m/(n-1)$.

Consider a given partition into t classes. At least half the number of classes in the partition have at most $2n/t$ nodes, twice the average number. Denote these classes as $SP = \{S_1, S_2, \ldots, S_{t'}\}$, where $t' \geq t/2$. The probability that S_i and S_j are not connected by an edge is

$$(1-p)^{|S_i|\cdot|S_j|} = e^{-p(2n/t)^2(1-o(1))} = e^{-2\log n/C^2} = n^{-1/C'},$$

for constant $C' = C^2/(2\log 2)$. The probability that SP is a complete partition is bounded above by

$$(1-n^{-1/C'})^{\binom{t}{2}} \leq e^{-m/(n^{1/C'}\log n)}.$$

The number of unordered partitions of vertices into t classes is

$$n^t = e^{t\ln n} = e^{m^{1/2+o(1)}}.$$

Thus, the probability that any of these partitions is complete is at most

$$e^{m^{1/2+o(1)} - m^{1-1/C'}},$$

which for $C' > 2$ is exponentially small. □

Acknowledgments

I thank Guy Kortsarz for discussions and for posing this problem, and I thank Jaikumar Radhakrishnan and Sivaramakrishnan Sivasubramanian for valuable help with the probabilistic argument for the case $\Delta(G) = n/\log n$.

References

[AP90] B. Awerbuch and D. Peleg. Sparse partitions. *Proc. IEEE FOCS*, pages 503–513, 1990.

[Bod89] H. L. Bodlaender. Achromatic number is NP-complete for cographs and interval graphs. *Inform. Process. Lett.*, 31(3):135–138, 1989.

[CE97] N. Cairnie and K. Edwards. Some results on the achromatic number. *J. Graph Theory*, 26(3):129–136, 1997.

[CE98] N. Cairnie and K. Edwards. The achromatic number of bounded degree trees. *Discrete Mathematics*, 188:87-97, 1998.

[FHKS00] U. Feige, M. Halldórsson, G. Kortsarz and A. Srinivasan. Approximating the domatic number. *SIAM J. on Computing* 32(1): 172–195, 2002.

[G69] R. P. Gupta. Bounds on the chromatic and achromatic numbers of complementary graphs. *In Recent Progress in Combinatorics* (Proceedings of Third Waterloo Conference on Combinatorics, Waterloo, 1968) (ed. W. T. Tutte), Academic Press, New York (1969), pp. 229-235.

[K92] J. Knisely, 1992. Cited in Stephen Hedetniemi's webpage on ”Open Problems in Combinatorial Optimization” at `http://www.cs.clemson.edu/~hedet/coloring.html`