Equipartitions of Graphs[*]

David Eppstein[†]
Joan Feigenbaum[‡]
Chung-Lun Li[§]

January 5, 1990

Abstract

Let G be an undirected graph on n nodes, and let k be an integer that divides n. A *k-equipartition* π of G is a partition of $V(G)$ into k equal-sized pieces $V_1, \ldots, V_k$. A pair V_i, V_j of distinct sets in π is called a *bad pair* if there is at least one edge v_i — v_j of $E(G)$ such that $v_i \in V_i$ and $v_j \in V_j$. The *parameterized equipartition problem* is: given G and k, find an optimal k-equipartition of G, i.e., one with the smallest possible number of bad pairs. More generally, a *nontrivial equipartition* of G is a k-equipartition, for some proper divisor k of n. The *equipartition problem* is: given G, find a nontrivial equipartition with the minimum number of bad pairs, where the minimum is taken over all divisors k of n and all k-equipartitions. We prove that there are relatively sparse graphs all of whose equipartitions have the maximum number of bad pairs (up to constant factors). We also prove that the parameterized and unparameterized versions of the equipartition problem are NP-hard.

Key words: Probabilistic Argument, NP-Completeness, Graph Partitioning, Graph Separators

---

[*]This paper was issued as an AT&T Bell Laboratories Technical Memorandum on August 31, 1989.

[†]Xerox PARC, 3333 Coyote Hill Road, Palo Alto, California 94304. Work done in part at Columbia University, supported by NSF grant CCR-88-14977.

[‡]AT&T Bell Laboratories, Room 2C473, 600 Mountain Avenue, Murray Hill, NJ 07974.

[§]Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027.

1 Introduction

Let G be an undirected graph on n nodes, and let k be an integer that divides n. A *k-equipartition* π of G is a partition of $V(G)$ into k equal-sized pieces V_1, …, V_k. A pair V_i, V_j of distinct sets in π is called a *bad pair* if there is at least one edge v_i — v_j of $E(G)$ such that $v_i \in V_i$ and $v_j \in V_j$. The *parameterized equipartition problem* is: given G and k, find an optimal k-equipartition π, i.e., one with the smallest possible number of bad pairs. We use $p(G, k)$ to denote the number of bad pairs in an optimal k-equipartition and call it the *k-equipartition number* of G.

More generally, an *equipartition* of G is a k-equipartition of G, for some k that divides n. A *nontrivial equipartition* is a k-equipartition in which $1 < k < n$. The equipartition problem is: given G, find a nontrivial equipartition of G with $\min_{k|n} p(G, k)$ bad pairs. We use $p(G)$ to denote $\min_{k|n} p(G, k)$ and call it the *equipartition number* of G.

The equipartition problem can be looked at as a hybrid of the classical graph-partitioning problem (see, e.g., the paper of Kernighan [6]) and the classical graph-separator problem (see, e.g., the survey of Chung [1]). In Section 2 below, we prove that there are relatively sparse graphs whose equipartition number is as high as possible. In Section 3, we prove the NP-hardness of both the parameterized and the unparameterized equipartition problems.

2 Density

Consider the question of how bad $p(G)$ can be. That is, are there relatively sparse graphs G with large equipartition number?

If $n = q^2$, where q is prime, then the total number of pairs in any nontrivial equipartition is $\binom{q}{2} = \Omega(n)$; hence graphs with equipartition number $\Omega(n)$ are, to within constant factors, as bad as possible.

Theorem 2.1 *There is an infinite family of graphs $\{G_n\}$ in which $|V(G_n)| = n = q^2$, where q is prime, $|E(G_n)| = O(n \log n)$, and $p(G_n) = \Omega(n)$.*

Proof:

We use a relatively straightforward random-graph argument. Let $n = q^2$, where q is prime, and fix a nontrivial equipartition $\pi = \{V_1, \ldots, V_q\}$ of V, a set of n nodes. Let G be a uniformly-distributed random element of the set of all labeled m-edge graphs on V. We estimate the probability that π is a "good" equipartition of G in the following sense.

Say that π is "bad for G" if at least 1/4 of the $\binom{q}{2}$ possible bad pairs are present; otherwise, say that π is "good for G." To prove the theorem, it suffices to show that, for all sufficiently large n, there is a G for which no good π exists.

Let $b = \lfloor 1/4\binom{q}{2} \rfloor$. Consider selecting a random G by sampling m edges $e_1, \ldots, e_m$ without replacement. If π is bad for G, then there is some $i \leq m$ for which e_i contributes the b^{th} bad pair. For $1 \leq j \leq i$, the probability that e_j contributes a new bad pair is greater than 1/2. Thus, the probability that fewer than b bad pairs are added during the m samples (i.e., the probability that π is good for G) is less than the probability that fewer than b heads occur in m fair coin tosses. The probability that some good partition exists for G is upper bounded by the product of the total number of partitions and the probability that a fixed partition is good, which in turn is less than

$$n^n b \binom{m}{b} 2^{-m}. \tag{1}$$

Now, $b < n/4$, by definition, and $m < n^2$, because it is the number of edges is an n-node graph. So (1) is upper bounded by

$$b \cdot n^n \cdot n^{n/2} \cdot 2^{-m} = 2^{(3n/2+1)\log n - m}.$$

As long as $m > (3n/2 + 1) \log n$, this probability is less than 1; hence, there is some G for which no good equipartition exists. ∎

Theorem 2.1 can be applied in the theory of lexicographic product graphs.[1] The *lexicographic product* $G[H]$ of graphs G and H has node set $V(G) \times V(H)$ and edge set $\{(u, v) — (u', v')\colon (u — v \in E(G))$ or $(u = v$ and $u' — v' \in E(H))\}$. More concretely, form the lexicographic product $G[H]$ by replacing each node in G with a *copy* of H and drawing in all possible edges between adjacent copies. (See Figure 1.) Graphs that cannot be written as lexicographic products in which both factors have two or more nodes are called *irreducible.* A *minimal factorable extension* of an irreducible graph G is a lexicographic

[1]For a thorough introduction to lexicographic product graphs and their applications in computer science, see, e.g., [2, 4].

product graph H for which three conditions hold: G is a subgraph of H; among all lexicographic product graphs of which G is a subgraph, H has the fewest nodes; among all lexicographic product graphs satisfying the first two conditions, H has the fewest edges. Note that, because $K_{qr} = K_q[K_r]$, a minimal factorable extension H of G has $|V(H)| = |V(G)|$ if $|V(G)|$ is composite and $|V(H)| = |V(G)| + 1$ if $|V(G)|$ is prime. Thus, the interesting question is how dense H can be in relationship to G.

Feigenbaum has observed that there is an infinite family $\{G_n\}$ of irreducible graphs with $|V(G_n)| = n = q^2$, where q is prime, and $|E(G_n)| = O(n^{3/2} \log n)$ for which $|V(H_n)| = \Omega(n^2)$; that is, the minimal factorable extensions H_n of G_n are as dense as possible [2]. The following corollary of Theorem 2.1 improves this observation.

Corollary 2.1 *There is an infinite family $\{G_n\}$ of irreducible graphs with $|V(G_n)| = n$ and $|E(G_n)| = O(n \log n)$ such that any minimal factorable extension H_n of G_n has $|E(H_n)| = \Omega(n^2)$.*

Proof: The family $\{G_n\}$ of Theorem 2.1 suffices. Any factorable extension H_n of G_n must be of the form $H_n = H_n^1[H_n^2]$, where $|V(H_n^1)| = |V(H_n^2)| = q$. The q disjoint copies of H_n^2 give a q-equipartition of H_n. By Theorem 2.1, any such equipartition has $\Omega(n)$ bad pairs, and hence any lexicographically factorable extension of G_n has $\Omega(n)$ adjacent copies. The contribution to $E(H_n)$ of edges that connect adjacent copies of H_n^2 is $\Omega(n \cdot q^2) = \Omega(n^2)$, because each pair of adjacent copies contributes q^2 edges. Thus the graphs G_n are irreducible and their minimal factorable extensions are as dense as possible. ∎

Equipartitions with $|E(G)| < |V(G)| \log |V(G)|$ have not been fully analyzed. In particular, we do not know tight bounds on the equipartition numbers of graphs with a linear number of edges. We have the following two partial results. The first is a generalization of Theorem 2.1.

Corollary 2.2 *Suppose that $1 \leq f(n) \leq \log n$. Then there is an infinite family of graphs $\{G_n\}$ in which $|V(G_n)| = n = q^2$, where q is prime, $|E(G_n)| = O((n \log n)/f(n))$, and $p(G_n) = \Omega(n/(f(n)^2))$. In particular, there is a family $\{G_n\}$ with $|E(G_n)| = O(n)$ and $p(G_n) = \Omega(n/(\log n)^2)$.*

Proof: Let $V = V(G_n)$ be a set of $n = q^2$ nodes; split V into two sets, X and Y, of sizes $n/f(n)$ and $n - n/f(n)$, respectively. Any q-equipartition π

of G_n induces a partition π' of X into q sets, some of which may be empty. (Note that distinct sets in π' are contained in distinct sets in π.) The sets in π' of size less than $q/2f(n)$ account for a total of at most $n/2f(n)$ nodes, leaving at least $n/2f(n)$ nodes of X in sets of size at least $q/2f(n)$. Each such larger set can only cover q of these nodes, so there must be at least $n/2qf(n) = q/2f(n)$ such sets.

Thus, for any equipartition of G_n, there is an induced equipartition π' of a subgraph V' of $V(G_n)$ consisting of $q/2f(n)$ pieces, each of size exactly $q/2f(n)$ (if necessary, remove nodes from some of the sets discussed above). Note that the probabilistic argument in Theorem 2.1 does not use the fact that the number (and size) of sets in the equipartition is prime. Apply the same argument, this time adding $(n/f(n))\log n$ random edges to V'; here it gives a π' for V' (and therefore also a π for V) with $O(n/(f(n)^2))$ bad pairs. ▮

Our other result on equipartitions of graphs with $|E(G)| = O(|V(G)|)$ shows that, in at least one important case, optimal equipartitions can be found in polynomial time.

Remark 2.1 *If $V(G_n) = n = q^2$, where q is prime, and G_n is planar, then $p(G_n) = O(\sqrt{n})$. There is a polynomial-time algorithm that finds a q-equipartition of G_n with $O(\sqrt{n})$ bad pairs.*

Proof: Let G_n be a planar graph. By a theorem of Yannakakis, G_n must have a four-page book embedding [8]. Fix such an embedding, and assume without loss of generality that the nodes of G_n appear on the spine of the book in the order $1, 2, \ldots, n$. Let $V_1, \ldots V_q$ be the q-equipartition of G_n in which $V_i = \{(i-1)q+1, \ldots, iq\}$. The induced graph with node-set $\{V_1, \ldots, V_q\}$ and edge-set $\{V_i — V_j\colon\ \exists\ v_i \in V_i$ and $v_j \in V_j$ with $v_i — v_j \in E(G_n)\}$ has an obvious four-page book embedding. Graphs with k pages and q nodes have $O(kq)$ edges. Hence, this q-equipartition of G_n shows that $p(G_n) = O(\sqrt{n})$. Furthermore, Yannakakis [8] gives a polynomial-time algorithm for finding a four-page book embedding of a planar graph, and the same algorithm can clearly be used here to find a q-equipartition with $O(\sqrt{n})$ bad pairs. Finally, note that the equipartition given by this algorithm is optimal (up to constant factors), because any q-equipartition of a connected graph on $n = q^2$ nodes has $\Omega(\sqrt{n})$ bad pairs. ▮

3 NP-Completeness

We now show that both the equipartition problem and the parameterized equipartition problem are NP-hard (as defined in Chapter 5 of Garey and Johnson [5]), which means that neither can be solved in polynomial time unless P = NP. To do this, it suffices to show that the following decision versions of the problems are NP-complete.

Parameterized Equipartition (PEP):
Input : A graph G and integers k and t.
Question : Is there a k-equipartition of G with at most t bad pairs?

Equipartition (EP):
Input : A graph G and an integer t.
Question : Is there an equipartition of G with at most t bad pairs?

Theorem 3.1 *The PEP problem is NP-complete.*

Proof: PEP is obviously in NP: given a k-equipartition V_1, …, V_k of G, it is easy to check in polynomial time that at most t pairs V_i, V_j have edges between them. The proof of completeness is by many-to-one reduction from the following problem.

3-Partition:
Input : A set $A = \{a_1, \ldots, a_{3n}\}$ of $3n$ distinct elements and an integer weight $s(a)$ for each $a \in A$, satisfying the conditions that $\sum_{a \in A} s(a) = nB$ and $B/4 < s(a) < B/2$, for all $a \in A$.
Question : Is there a partition of A into disjoint sets $\{a_{i0}, a_{i1}, a_{i2}\}$, $1 \leq i \leq n$, such that $s(a_{i0}) + s(a_{i1}) + s(a_{i2}) = B$, for each i?

The 3-Partition problem is NP-complete, even if the numbers $s(a_i)$ and B are written in unary [5].

Let $(A = \{a_1, \ldots, a_{3n}\}, s, B)$ be an instance of 3-Partition. We construct an equivalent instance of PEP as follows. For each element $a_i \in A$, let $C_i \cong K_{s(a_i)}$ be a clique on $s(a_i)$ nodes. Let $C_0 \cong K_B$ be a clique on B nodes. The node set of the graph G in the PEP instance is just the disjoint union of the node sets $V(C_i)$, $0 \leq i \leq 3n$. The edge set of G consists of the disjoint union of the edge sets $E(C_i)$, $0 \leq i \leq 3n$, together with all possible edges v — w such that $v \in V(C_0)$ and $w \notin V(C_0)$. (See Figure 2.) Finally, let $k = n+1$ and $t = n$. So (G, k, t) is a yes-instance of PEP if and only if $V(G)$

can be partitioned into $n + 1$ sets, each containing B nodes, such that the number of bad pairs of sets is at most n. It should now be clear why it is important that the numbers in the 3-Partition instance be written in unary; otherwise this reduction would entail an exponential blow-up in the size of the instance.

If (A, s, B) is a yes-instance of 3-Partition, then $(G, n + 1, n)$ is clearly a yes-instance of PEP: if $\{a_{10}, a_{11}, a_{12}\}$, ..., $\{a_{n0}, a_{n1}, a_{n2}\}$ is a partition of A that witnesses the fact that (A, s, B) is a yes-instance, then $V(C_0)$, $V(C_{10}) \cup V(C_{11}) \cup V(C_{12})$, ..., $V(C_{n0}) \cup V(C_{n1}) \cup V(C_{n2})$ is an $(n+1)$-equipartition of $V(G)$ that witnesses the fact that $(G, n + 1, n)$ is a yes-instance.

Conversely, suppose that $(G, n + 1, n)$ is a yes-instance of PEP and that V_0, ..., V_n is an $(n+1)$-equipartition of $V(G)$ that witnesses this fact. Let v and w be distinct nodes in the large clique C_0. Then v and w must be in the same set V_i. Suppose, to the contrary, that $v \in V_0$ and $w \in V_1$. Then V_0, V_1 is a bad pair, V_0, V_i, is a bad pair, and V_1, V_i is a bad pair, for all $2 \leq i \leq n$. Thus, the number of bad pairs is at least $2n - 1$, which is greater than n and thus contradicts the fact that $(G, n + 1, n)$ is a yes-instance.

So we can assume without loss of generality that $V_0 = V(C_0)$. The nodes of the remaining $3n$ cliques in $V(G)$ must be partitioned into V_1, ..., V_n. Any such equipartition causes each of the n pairs V_0, V_i, $1 \leq i \leq n$ to be bad, because each node in C_0 is adjacent to each node in every other clique C_i. Thus none of the pairs V_i, V_j, $1 \leq i < j \leq n$, can be bad; this means that each node set $V(C_i)$, $1 \leq i \leq n$, is a subset of some V_j in the equipartition. No four such node sets can be in the same V_j, because $|V(C_i)| = s(a_i) > B/4$, and $|V_j| = B$. The total number of sets $V(C_i)$ is $3n$; hence, each V_j contains exactly three of them. If $V_j = V(C_{j0}) \cup V(C_{j1}) \cup V(C_{j2})$, $1 \leq j \leq n$, then the corresponding partition $\{a_{j0}, a_{j1}, a_{j2}\}$, $1 \leq j \leq n$ of A satisfies

$$\sum_{p=0}^{2} s(a_{jp}) = \sum_{p=0}^{2} |V(C_{jp})| = |V_j| = B,$$

for each j and thus witnesses the fact that (A, s, B) is a yes-instance of 3-Partition. ▮

Theorem 3.2 *The EP problem is NP-complete.*

Proof: Once again, it is obvious that EP is in NP. We reduce 3-Partition to EP. Let (A, s, B) be a 3-Partition instance. Let r be the smallest prime number such that $r \geq n$ and R be the smallest prime number such that $R > (r+1)B$. By a theorem of Chebyshev [7], we know that $r < 2n$ and that $R < 2((r+1)B+1)$; hence, both can be found by trial division of successive odd integers in time polynomial in nB. (Similar uses of Chebyshev's theorem in graph-theoretic problems can be found in [2, 3, 4].) The graph G of the target EP instance is the disjoint union of $2r+2n$ complete graphs: one copy of $K_{s(a_i)}$, for $1 \leq i \leq 3n$, $a_i \in A$, $r-n$ copies of K_B, and r copies of K_{R-B}. (See Figure 3.) The parameter t of the EP instance can be taken to be 0.

If (A, s, B) is a yes-instance of 3-Partition, then $(G, 0)$ is easily seen to be a yes-instance of EP; the r-equipartition of G with no bad pairs is shown in Figure 4.

Conversely, suppose that G has a k-equipartition with no bad pairs. Because $|V(G)| = rR$, and both r and R are prime, either $k = r$ or $k = R$. By construction, $R - B > r$, and K_{R-B} is a subgraph of G. Thus $V(G)$ cannot be partitioned into R disjoint sets of size r with no bad pairs. Hence $k = r$.

Each of the copies of K_{R-B} must lie completely within one set in the r-equipartition. By construction, $2(R-B) > R$, and so no two copies of K_{R-B} can be in the same set. This implies that $r-n$ of the sets in the equipartition each consist of one copy of K_{R-B} and one copy of K_B. Furthermore, the remaining n sets each contain one copy of K_{R-B} and cliques of the form $K_{s(a_i)}$; again, the entire clique $K_{s(a_i)}$ must be contained in a single set of the equipartition if there are to be no bad pairs. In each set, the number of nodes contributed by cliques of the form $K_{s(a_i)}$ is n. Because $B/4 < s(a_i) < B/2$, there must be exactly three such cliques in each set. Thus, they determine a solution to the 3-Partition instance.

Finally, observe that, as in Theorem 3.1, it is important that the input be given in unary for this reduction to be polynomial-time. ∎

Theorem 3.1 exhibits a many-to-one reduction from 3-Partition to PEP. The fact that PEP is complete for NP under Turing reductions follows from Theorem 3.2. The instance (G, t) of EP is a yes-instance if and only if (G, k, t) is a yes-instance of PEP, for some nontrivial divisor k of $|V(G)|$. All of the nontrivial divisors of $|V(G)|$ can be found by trial division in time polynomial in $|V(G)|$; hence there is an obvious Turing reduction from EP to PEP.

If we relax the definitions of the equipartition problems so that the sets in the partition are allowed to differ in size by one, then EP and PEP are apparently different in complexity. Suppose that a solution to the PEP instance (G, k, t) were allowed to have j sets of size $\lfloor |V(G)|/k \rfloor$ and $k - j$ sets of size $\lceil |V(G)|/k \rceil$. Then, this apparently easier problem would still be NP-complete, because it contains the original PEP problem, in which all yes-instances satisfy $k \mid |V(G)|$, as a special case. In the EP problem instance (G, t), on the other hand, if the sets in a solution were allowed to differ in size by one, then every instance in which $t \geq 1$ would be a yes instance: we could partition $V(G)$ into two sets, one of size $\lfloor |V(G)|/2 \rfloor$ and one of size $\lceil |V(G)|/2 \rceil$, and have only one bad pair.

4 Acknowledgements

We are grateful to Fan Chung for suggesting that random graphs could be used for this problem and to Larry Ozarow for helpful discussions.

References

[1] F. R. K. Chung, *Separator Theorems with Applications*, to appear.

[2] J. Feigenbaum, *Lexicographically Factorable Extensions of Irreducible Graphs*, to appear in Proceedings of the 6th International Conference on the Theory and Applications of Graphs, Kalamazoo, MI, May, 1988, John Wiley and Sons.

[3] J. Feigenbaum and R. W. Haddad, *On Factorable Extensions and Subgraphs of Prime Graphs*, SIAM J. Disc. Math. (2), 1989, 197–218.

[4] J. Feigenbaum and A. A. Schäffer, *Recognizing Composite Graphs is Equivalent to Testing Graph Isomorphism*, SIAM J. Comput. (15), 1986, 619–627.

[5] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman and Company, San Francisco, 1979.

[6] B. W. Kernighan, *Optimal Sequential Partitions of Graphs*, J. Assoc. Comput. Mach. (18), 1971, 34–40.

[7] I. Niven and H. S. Zuckerman, *An Introduction to the Theory of Numbers*, 3rd ed., John Wiley and Sons, New York, 1972.

[8] M. Yannakakis, *Embedding Planar Graphs in Four Pages*, J. Comput. Sys. Sci. (38), 1989, 36–67.

Figure 1: The lexicographic product of a three-node chain and a triangle.

Figure 2: PEP instance corresponding to the 3-Partition instance
$(s(a_1) = s(a_2) = s(a_3) = s(a_4) = 1, s(a_5) = s(a_6) = 2, B = 4)$,
as in the proof of Theorem 3.1.

Figure 3: EP instance corresponding to the 3-Partition instance $(s(a_1) = s(a_2) = s(a_3) = s(a_4) = 1, s(a_5) = s(a_6) = 2, B = 4)$, as in the proof of Theorem 3.2.

Figure 4: Witness that the target instance is in EP if the domain instance is in 3-Partition, as in the proof of Theorem 3.2.