Long Local Searches for Maximal Bipartite Subgraphs

Hemanshu Kaul* and Douglas B. West†

October 30, 2007

Abstract

Given a partition of the vertices of a graph into two sets, a *flip* is a move of a vertex from its own set to the other, under the condition that it has more incident edges to vertices in its own set than in the other. Every sequence of flips eventually produces a bipartite subgraph capturing more than half of the edges in the graph. Each flip gains at least one edge. For an n-vertex loopless multigraph, we show that there is always a sequence of at most $n/2$ flips that cannot be extended, and we construct a graph having a sequence of $\frac{2}{25}(n^2 + n - 31)$ flips.

1 Introduction

Finding a bipartite subgraph with the maximum number of edges in a given graph is a classical problem in combinatorial optimization and extremal graph theory. It has been studied in computer science from an algorithmic point of view (as the MAX-CUT problem) and in combinatorics from a more structural point of view. The problem is NP-complete (Karp [11]), so the focus of research has been on heuristics, approximation algorithms, and extremal results (conditions on graph parameters that guarantee large bipartite subgraphs).

For a graph G, let $b(G)$ denote the maximum number of edges in a bipartite subgraph of G. For the graph G in which we study $b(G)$, we always let n and m denote the numbers of vertices and edges. Classical guarantees include $b(G) \geq \frac{m}{2} + \frac{1}{8}(\sqrt{8m+1} - 1)$ (Edwards [5]) and, for connected graphs, $b(G) \geq \frac{m}{2} + \frac{1}{4}(n-1)$ (conjectured by Erdős [7] and proved

*Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL 60616, *kaul@math.iit.edu*

†Department of Mathematics, University of Illinois, Urbana, IL 61801, *west@math.uiuc.edu*. Supported in part by NSA grant H98230-06-1-0065.

by Edwards [6]). Bylka, Idzik, and Tuza [3] strengthened the Edwards–Erdős bound to $b(G) \geq \left\lceil \frac{m}{2} + \frac{1}{2} \left\lceil \frac{2t-1}{2t}(n-1) \right\rceil \right\rceil$ when G is connected and has no odd cycle of length less than $2t$. Furthermore, this is sharp (using graphs whose blocks are $(2t+1)$-cycles), and there is a linear-time algorithm to find a bipartite subgraph this big.

Bollobás and Scott [1] gave an algorithm running in time $O(2^{ck^4} + m + n)$ that finds a bipartite subgraph with at least $\frac{m}{2} + \sqrt{\frac{m}{8}} + k$ edges, if one exists, and otherwise provides an optimal partition (they also provide a survey of work in the area, updating the survey [14]). Goemans and Williamson ([9]) gave a randomized .87-approximation algorithm for $b(G)$ based on semidefinite programming. However, Håstad [10] proved that no .942-approximation algorithm exists unless P=NP.

An elementary heuristic approach to finding large bipartite subgraphs is to start with an arbitrary vertex partition and then make local improvements. Bylka, Idzik, and Tuza [3] found a local switching algorithm that guarantees a cut as big as the bound of Edwards [5]. A simple type of local improvement is to move a vertex having more incident edges to vertices in its own set than in the other set; moving it to the other set increases the size of the cut. We call such a change a *flip* of the vertex. Erdős [8] introduced this notion to give a short proof that every graph G has a bipartite subgraph with at least half its edges. When flips are no longer possible, for each vertex at least half of the incident edges appear in the current bipartite subgraph, and hence by summing degrees the subgraph contains at least half the edges of G.

Since the process also must terminate (in at most m flips), this simple process is a .5-approximation algorithm for $b(G)$. Furthermore, it lends itself to parallel computation, since the determination of whether a vertex can flip is independent of that determination for other vertices. Parallelizable local-improvement algorithms based on switching protocols have been studied by Bylka, Idzik, and Tuza [3], by Bylka [2], and others.

Our purpose in this paper is to study the *length* (number of flips) of a maximal sequence of flips, where *maximal* means that at the end every vertex has at least as many neighbors in the opposite side of the cut as in its own side. At that point, the cut is a local optimum for the MAX-CUT problem.

For the complete graph K_n, every maximal flip sequence has length $\lfloor n/2 \rfloor$. One might think that there is always a linear bound on the length of flip sequences, but Cowen and West [4] showed that this is false. They posed several questions about the lengths of maximal flip sequences in a graph G (we can consider these also for loopless multigraphs).

1) Let $l(G)$ denote the maximum length of a flip sequence in G starting from the trivial vertex partition $(\varnothing, V(G))$. What is the maximum of $l(G)$ over n-vertex graphs? Over n-vertex graphs with m edges?

2) Let $s(G)$ denote the minimum length of a maximal flip sequence starting from the trivial vertex partition. What is the maximum of $s(G)$ over n-vertex graphs? Is it true that always $s(G) \le n/2$, and is there a fast algorithm to produce a flip sequence of length at most $n/2$?

Since every flip increases the number of edges across the cut, $l(G) \le m$. When n is a perfect square and $m = \frac{1}{2}(n-1)\sqrt{n}$, Cowen and West [4] constructed a graph with a flip sequence of length m, matching the edge bound. It remains open whether $\frac{1}{2}n^{3/2}$ is asymptotically the maximum number of edges in a graph with a flip sequence whose length is the number of edges (increasing the size of the cut by only 1 with each flip).

The construction with length $\frac{1}{2}n^{3/2}$ was at that time the longest known flip sequence in an n-vertex graph. Since each flip increases the size of the cut, $\frac{1}{4}n^2$ is a trivial upper bound. Our main result, Theorem 3, is construction of an n-vertex graph having a flip sequence of length at least $\frac{2}{25}(n^2+n-31)$. This graph has roughly $\frac{1}{5}n^2$ edges.

Flips can be found in time proportional to maximum degree, so $l(G)$ times maximum degree is a bound on the worst-case running time of this simple heuristic algorithm. See Korte and Lovász [12], Loebl [13], and Tovey [15] for further discussion of issues related to complexity of greedy and local-search based algorithms. Our result shows that one cannot guarantee less than $O(n^2)$ for the worst-case number of steps.

It would be interesting to know the coefficient on n^2 in the largest $l(G)$. We believe our construction is near optimal. In Proposition 2, we show that the maximum length of a flip sequence in a graph G with minimum degree d is at most $b(G) - \lfloor d^2/4 \rfloor$.

We also answer most of question (2), proving in Theorem 1 that $s(G) \le \lfloor n/2 \rfloor$ for every n-vertex loopless multigraph G. The bound is sharp, as every maximal flip sequence in K_n has length $\lfloor n/2 \rfloor$. It is also sharp for $K_{n/2,n/2}$, where every local optimum splits the vertices in half (including the "worst" local optimum with $n/4$ vertices of each partite set on each side when 4 divides n). Our proof does not provide a fast algorithm to find such a sequence, since it assumes knowledge of an optimal partition (in fact, it reaches some optimal partition in at most $\lfloor n/2 \rfloor$ flips). Thus the complexity of computing $s(G)$ or of finding a flip sequence of length at most $\lfloor n/2 \rfloor$ remains open. Is there a polynomial-time algorithm to find a sequence of $O(n)$ flips that result in a maximal bipartite subgraph? In terms of parallel computation, is there an NC algorithm that will construct a maximal bipartite subgraph?

There remain other open questions about local switching algorithms for MAX-CUT. Bylka, Idzik, and Tuza [3] showed that no algorithm using flips and certain other local switches guarantees a cut as big as the Edwards–Erdős bound [6]. A more general algorithm might come closer to that bound. Fix a positive integer k. Given a current vertex partition, a *local k-switch* selects subsets of size at most k from the two parts and switches them; it is *improving* if the new cut is bigger than the old one. Tuza [14] asked for the largest constant $c(k)$ such that a sequence of improving local k-switches can produce a bipartite subgraph with $\frac{m}{2} + c(k)n - o(n)$ edges.

2 Upper Bound on $s(G)$

We use $N(v)$ to denote the neighborhood of a vertex v (in G). For $S \subseteq V(G)$ and $v \in V(G)$, let $d_S(v)$ be the number of edges incident to v whose other endpoints are in S. Recall that $s(G)$ is the minimum length of a maximal flip sequence from the trivial partition.

Theorem 1 *If G is an n-vertex loopless multigraph, then $s(G) \le \lfloor n/2 \rfloor$. In particular, given a globally optimal partition, a sequence of at most $\lfloor n/2 \rfloor$ flips can be constructed that produces some globally optimal partition.*

Proof. Let (X, Y) be a globally optimal partition, with $|X| = k \le n/2$. By starting with $(\varnothing, V(G))$ and flipping to the first part some or all of the members of the set X, we will obtain a globally optimal partition $(\hat{X}, \hat{Y})$ with $\hat{X} \subseteq X$ and $\hat{Y} \supseteq Y$.

Let X' denote the set of vertices in X having exactly half their neighbors in X and half in Y. Let $G_1, \ldots, G_r$ be the components of the induced subgraph $G[X']$. Select $x_i \in V(G_i)$, for $1 \le i \le r$. Let $Z = \{x_1, \ldots, x_r\}$. Index the rest of X as $x_{r+1}, \ldots, x_k$ in such a way that each vertex of $X' - Z$ has an earlier neighbor in X'. In particular, choose a spanning tree of each G_i, root it at x_i, and search the tree from x_i, assigning labels in increasing order as vertices are found.

Now flip the vertices $x_k, \ldots, x_{r+1}$ (from the second part to the first), in decreasing order of index. We claim that these flips are all available when desired and that the resulting partition $(\hat{X}, \hat{Y})$ is globally optimal. Since (X, Y) is globally optimal, $d_X(u) \le d_Y(u)$ whenever $u \in X$, and $d_Y(v) \le d_X(v)$ whenever $v \in Y$. Furthermore, these inequalities are strict for all vertices of odd degree.

With the given ordering of X, let $X_i = \{x_{k+1-i}, \ldots, x_k\}$ for $1 \le i \le k - r$, with $X_0 = \varnothing$. Let $Y_i = V(G) - X_i$ for $0 \le i \le k - r$. Since $X_{i-1} \subset X$ and $Y \subset Y_{i-1}$, we have $d_{X_{i-1}}(x_i) \le d_X(x_i) < d_Y(x_i) \le d_{Y_{i-1}}(x_i)$ when $x_i \notin X'$. For $x_i \in X' - Z$, the indexing gives x_i a neighbor in $X \cap Y_{i-1}$. Thus the inequalities become $d_{X_{i-1}}(x_i) < d_X(x_i) \le d_Y(x_i) < d_{Y_{i-1}}(x_i)$. The two cases allow us to flip $x_k, \ldots, x_{r+1}$ in order.

We have reached a partition $(\hat{X}, \hat{Y})$ with $\hat{X} = X - Z$ and $\hat{Y} = Y \cup Z$. It remains to show that $(\hat{X}, \hat{Y})$ is a global optimum, which also implies that no further flips can be made. Reaching (X, Y) from $(\hat{X}, \hat{Y})$ is accomplished by moving the vertices of Z. Because Z consists of one vertex from each component of $G[X']$, Z is an independent set. Since also $Z \subseteq X'$, $d_{\hat{X}}(x) = d_{\hat{Y}}(x)$ for each $x \in Z$. Therefore, moving all of Z does not change the size of the cut, and $(\hat{X}, \hat{Y})$ is already optimal. $\square$

If every vertex of G has odd degree, then the first statement of Theorem 1 can be proved more simply. In this case, we do not need to know a globally optimal partition to obtain a maximal flip sequence of length at most $\lfloor n/2 \rfloor$. We can start with any locally optimal partition, which can be found in polynomial time (as discussed in the introduction), and this partition can be reached from the trivial partition simply by flipping each vertex of one part. When G has vertices of even degree, our proof uses a given globally optimal partition to prevent vertices of Y from flipping; finding such a partition is NP-hard.

3 Upper bound on $l(G)$

Henceforth we consider only simple graphs. Recall that $l(G)$ is the maximum length of a flip sequence from the trivial partition of G. Since each flip increases the number of edges across the cut, $l(G) \le b(G)$, and it is well known that $b(G) \le n^2/4$. The example constructed by Cowen and West [4] is bipartite, with $l(G) = b(G) = m$. Achieving length $b(G)$ requires gaining only one edge with each flip, so G must have minimum degree 1. Their construction uses about $\sqrt{n}$ vertices of degree 1. It is natural to wonder how by much the upper bound $b(G)$ can be lowered when G has minimum degree d.

The basic idea is that the first few switches must gain many edges, leaving fewer for later flips. It may be happen that the later flips gain only one edge each. Indeed, this is what happens in our construction, where the minimum degree is linear in n and yet the number of flips is quadratic.

Proposition 2 *If a graph G has minimum degree d, then $l(G) \le b(G) - \lfloor d^2/4 \rfloor$.*

Proof. In a flip sequence on G from the initial partition with $(X, Y) = (\varnothing, V(G))$, the first move gains at least d edges. The second flip must move another vertex to X. As long as at most $1 + d/2$ vertices have been flipped, no vertex of X can have more neighbors in X than in Y, so no vertex in X can be flipped back to Y.

When $k = \lfloor d/2 \rfloor$, the first k flips thus yield a partition with $|X| = k$; at most $\binom{k}{2}$ edges are induced by X. Hence at this point the cut has at least $kd - 2\binom{k}{2}$ edges. Each subsequent flip after the first k gains at least one edge in the cut, until at most $b(G)$ edges have been captured. The total number of flips is thus at most $k + b(G) - (kd - 2\binom{k}{2})$, which simplifies to $b(G) - k(d-k)$. With $k = \lfloor d/2 \rfloor$, the bound follows. □

To see the effect of this bound, suppose that $4 \nmid n$ and $b(G)$ is $n^2/4$, the maximum possible value. This requires G to have the complete bipartite graph $K_{n/2,n/2}$ as a subgraph, and hence G has minimum degree at least $n/2$. Now Proposition 2 yields $l(G) \le \frac{n^2}{4} - \frac{1}{4}\frac{n^2}{4} = \frac{3}{16}n^2$. More generally, if G is αn-regular, where $0 < \alpha \le 1$, using $b(G) \le m$ yields $l(G) \le m - \frac{1}{4}\alpha^2 n^2 = m(1 - \frac{1}{2}\alpha)$.

For the graph G we construct in Theorem 3, where $l(G) > \frac{2}{25}n^2$, we have $b(G) \approx \frac{4}{25}n^2$, and the minimum degree is about $\frac{2}{5}n$. The resulting upper bound on $l(G)$ from Proposition 2 is about $\frac{3}{25}n^2$, which equals $\frac{3}{2}(\frac{2}{25}n^2)$. This suggests that it will be hard to improve Proposition 2 without knowing something about the structure of G. For example, when G is triangle-free, the gain in the first k flips in the proof of Proposition 2 is at least $kd - 2k^2/4$. This improves the upper bound in Proposition 2 to roughly $b(G) - \frac{3}{8}d^2$, and it improves the upper bound on $l(G)$ for the graph in Theorem 3 to about $\frac{5}{4}(\frac{2}{25}n^2)$.

4 Lower Bound Construction for $l(G)$

In this section, we construct a graph G and a flip sequence of length about $\frac{2}{25}n^2$ on it. As mentioned earlier, Cowen and West [4] achieved a sequence of length $\frac{1}{2}n^{3/2}$ using a graph with that many edges and a sequence that increases the size of the cut by 1 with each flip. We achieve quadratic length by sacrificing equality with m and enlarging the cut rapidly at first in order to permit many more small steps later in which vertices repeatedly move back and forth. With such a sequence we come near the counting bound from Proposition 2.

Theorem 3 *For every positive integer n, there exists an n-vertex graph G such that $l(G) \ge \frac{2}{25}(n^2 + n - 31)$.*

Proof. We first present an explicit construction for $n \equiv 3 \mod 5$, with $l(G) > \frac{2}{25}n^2$. Given a positive integer k, let $n = 5k + 3$. Create vertex subsets $V_1, \ldots, V_5$, with $|V_1| = |V_2| = k$ and $|V_3| = |V_4| = |V_5| = k+1$. Form G by expanding the vertices of a 5-cycle into these sets. That is, given the 5-cycle with vertices $z_1, \ldots, z_5$, replace z_i with V_i for each i, with every vertex in V_i adjacent to every vertex in V_{i-1} and V_{i+1}, treating subscripts modulo 5. Label the vertices in the first three sets as follows: $V_1 = \{w_1, \ldots, w_k\}$, $V_2 = \{u_1, \ldots, u_k\}$, and $V_3 = \{v_1, \ldots, v_{k+1}\}$. See Figure 1.

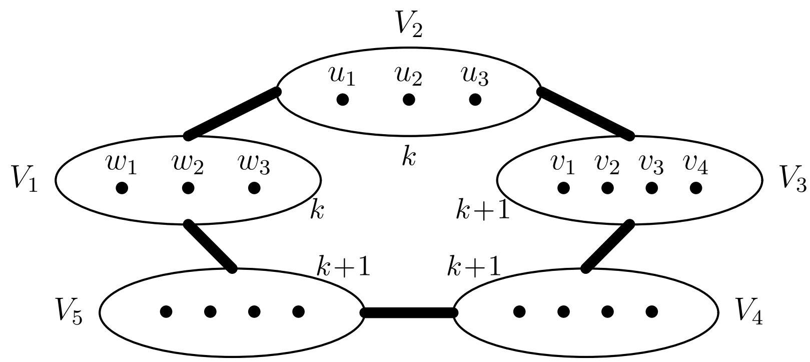

Figure 1: Graph G in Theorem 3, with $k = 3$.

We start with the trivial partition $(V(G), \varnothing)$. During the flip sequence, we use X and Y respectively to denote the first and the second set in the current partition, so initially $X = V(G)$. At the end of the flip sequence, we will have $X = V_1 \cup V_2 \cup V_4$ and $Y = V_3 \cup V_5$. As explained after the proof, this in fact is an optimal cut; it has $2k(k+1) + 2(k+1)^2$ edges, including all the edges of G except the k^2 edges between V_1 and V_2.

Our flip sequence for G begins with a "preprocessing" Phase 0 that gains many edges for each flip. Since G has large minimum degree, the argument in Proposition 2 implies that the flip sequence must start this way.

Phase 0 consists of two steps. We first flip each vertex of V_1 from X to Y. This is valid because all neighbors of each vertex of V_1 lie in $X - V_1$ throughout this phase. We next flip each vertex of V_5 from X to Y. This is valid because, after moving V_1, each vertex of V_5 has $k+1$ neighbors in X (all of V_4) and k neighbors in Y (all of V_1) throughout the remainder of this phase.

After Phase 0, we have $V_4 \subset X$ and $V_5 \subset Y$, and this remains true for the remainder of the flip sequence. We group the remainder of the flips into Phases 1 through $k+1$.

Let $V_1(i) = \{w_1, \ldots, w_{i-1}\}$ and $V_1'(i) = \{w_i, \ldots, w_k\}$. Similarly, let $V_3(i) = \{v_i, \ldots, v_{i-1}\}$ and $V_3'(i) = \{v_i, \ldots, v_{k+1}\}$. Thus $V_1 = V_1'(1)$ and $V_3 = V_3'(1)$ and $V_1(1) = V_3(1) = \varnothing$. Hence the cut at the end of Phase 0 can be described as $X = V_1(1) \cup V_2 \cup V_3'(1) \cup V_4$ and $Y = V_1'(1) \cup V_3(1) \cup V_5$, which is the description of the starting cut for Phase 1. The four steps of Phase i are shown in Figure 2.

Phase i: At the start of Phase i, $X = V_1(i) \cup V_2 \cup V_3'(i) \cup V_4$ and $Y = V_1'(i) \cup V_3(i) \cup V_5$.
Step a_i: Flip all of V_2 from X to Y, valid because each vertex in V_2 now has $k+1$ neighbors in X (all of $V_1(i) \cup V_3'(i)$) and k neighbors in Y (all of $V_1'(i) \cup V_3(i)$).
Step b_i: Flip v_i in V_3 from X to Y, valid because v_i now has $k+1$ neighbors in X (all of

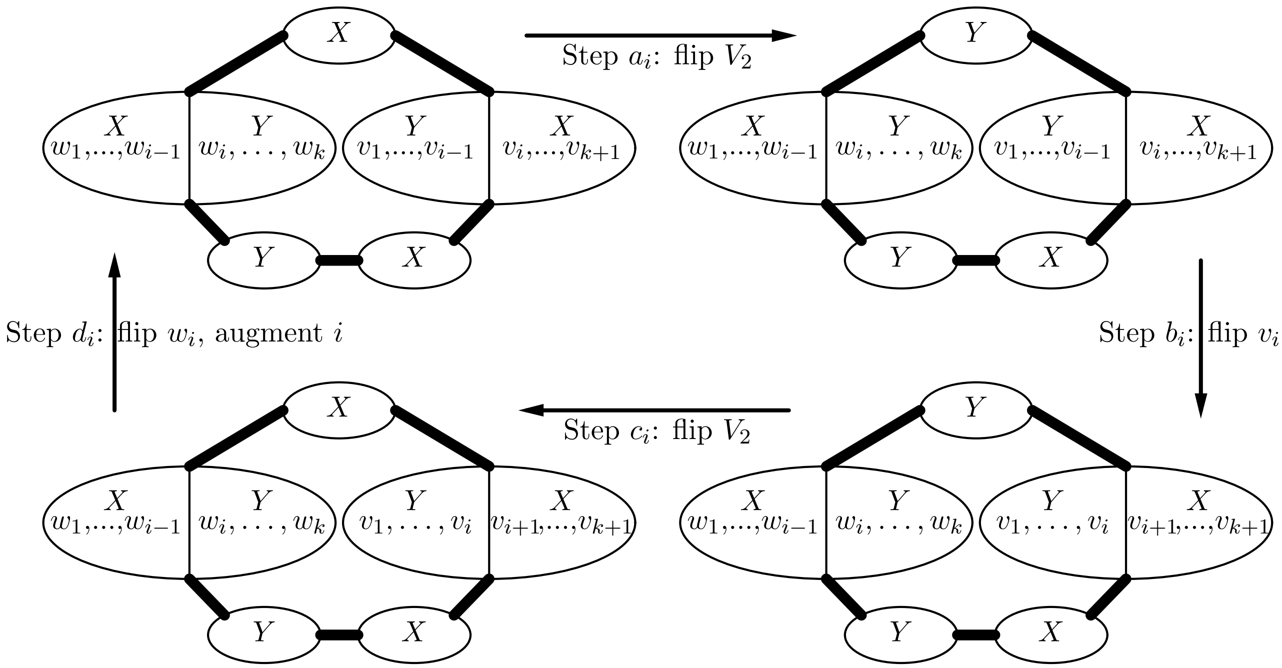

Figure 2: Steps in Phase i

V_4) and k neighbors in Y (all of V_2).
Step c_i: Flip all of V_2 from Y to X, valid because each vertex in V_2 now has $k+1$ neighbors in Y (all of $V_3(i+1) \cup V_1'(i)$) and k neighbors in Y (all of $V_3(i) \cup V_1'(i)$).
Step d_i: If $i \leq k$, flip w_i in V_1 from Y to X; otherwise stop. For $i \leq k$, the flip is valid because w_i now has $k+1$ neighbors in Y (all of V_5) and k neighbors in X (all of V_2).

At the end of Phase $k+1$, the partition is $X = V_1 \cup V_2 \cup V_4$ and $Y = V_3 \cup V_5$. The flip sequence ends, since each vertex having any neighbors in its own side of the cut has k neighbors in its side and $k+1$ in the opposite side.

In Phase 0, both V_1 and V_5 were flipped; altogether $2k+1$ flips. Each subsequent Phase flips V_2 twice and flips one vertex of V_3 and one vertex of V_1, except that the last phase omits the last flip. Thus the total number of flips is $2k + (k+1)(2k+2)$. Since $n = 5k+3$, we obtain $l(G) \geq \frac{2}{25}(n^2 + 9n - 11)$.

Finally, suppose that $n \not\equiv 3 \mod 5$. Let n' be the largest integer congruent to 3 modulo 5 that is at most n. We use an instance of the construction above with n' vertices and leave the remaining vertices isolated. Since $n' \geq n-4$, we have $l(G) \geq \frac{2}{25}(n'^2 + 9n' - 11) \geq \frac{2}{25}(n^2 + n - 31)$. □

In fact, the final partition in the flip sequence we constructed in the proof of Theorem 3 is globally optimal, as we now show. Let G be any expansion of a 5-cycle replacing the ith vertex with an independent V_i, indexed cyclically. For any partition (X, Y) of $V(G)$ yielding a bipartite subgraph H, let $a_i = |X \cap V_i|$ and $b_i = |Y \cap V_i|$. Each vertex v in V_i has $a_{i-1} + a_{i+1}$ neighbors in X and $b_{i-1} + b_{i+1}$ neighbors in Y. Vertex v has largest degree in H if it is in the set opposite the larger of these two quantities. This applies for all $v \in V_i$ (and for all i), so we achieve the largest bipartite subgraph by putting each V_i onto one side of the cut. Under this restriction, the best we can achieve is to capture all but the smallest of the five complete bipartite subgraphs expanding edges of the cycle, which is what the final partition of Theorem 3 does.

References

[1] B. Bollobás, and A. D. Scott, Better bounds for Max Cut, in *Contemporary combinatorics.* Bolyai Soc. Math. Stud. 10, 185–246. János Bolyai Math. Soc., Budapest, 2002.

[2] S. Bylka, Local improving algorithms for large cuts in graphs with maximum degree three. *Discrete Mathematics* 235 (2001), 53–67.

[3] S. Bylka, A. Idzik, and Z Tuza, Maximum cuts - Improvements and local algorithmic analogues of the Edwards-Erdős inequality. *Discrete Mathematics* 194 (1999), 39–58,

[4] L. J. Cowen and D. B. West, Greedy vertex partition of a graph, Problem 382, in Research Problems, *Kleitman and Combinatorics: A Celebration Discrete Mathematics* 257 (2002), 599-624.

[5] C. S. Edwards, Some extremal properties of bipartite subgraphs. *Canad. J. Math.* 25 (1973), 475–485.

[6] C. S. Edwards, An improved lower bound for the number of edges in a largest bipartite subgraph. In *Recent advances in graph theory (Proc. Second Czechoslovak Sympos., Prague, 1974)*, (Academia, Prague, 1975), 167–181.

[7] P. Erdős, On some extremal problems in graph theory. *Israel J. Math.* 3 (1965), 113–116.

[8] P. Erdős, Problems and results on finite and infinite graphs. In *Recent advances in graph theory (Proc. Second Czechoslovak Sympos., Prague, 1974)*, (Academia, Prague, 1975), 183–192.

[9] M. X. Goemans and D. P. Williamson, Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming, *J. Assoc. Comp. Mach.* 42 (1995), 1115–1145.

[10] J. Håstad, Some Optimal inapproximability results. *Proc 29th Annual ACM Symp. Th. Comput., El Paso, 1997*, 1-10; *J. Assoc. Comp. Mach.* 48 (2001), 798–859.

[11] R. M. Karp, Reducibility among combinatorial problems. In *Complexity of computer computations (Proc. Sympos., IBM Thomas J. Watson Res. Center, Yorktown Heights, N.Y., 1972)*, (Plenum, New York, 1972), 85–103.

[12] B. Korte and L. Lovász, Mathematical Structures Underlying Greedy Algorithms. In *Lecture Notes In Computer Science*, 117, (Springer, Berlin, 1981), 205–209.

[13] M. Loebl, Efficient Maximal Cubic Graph Cuts. In *Lecture Notes in Computer Science*, 510, (Springer, Berlin, 1991), 351–362.

[14] S. Poljak and Z. Tuza, Maximum cuts and large bipartite subgraphs. *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.* 20 (1995), 181–244.

[15] C. Tovey, Local improvement on discrete structures. In *E. Aarts, J.K. Lenstra (eds.), Local Search in Combinatorial Optimization*, Wiley-Interscience Series, Wiley, New York, 1997.