

# Improving the Smoothed Complexity of FLIP for Max Cut Problems

ALI BIBAK, University of Illinois, Urbana-Champaign

CHARLES CARLSON, University of Colorado Boulder

KARTHEKEYAN CHANDRASEKARAN, University of Illinois, Urbana-Champaign

Finding locally optimal solutions for MAX-CUT and MAX- $k$ -CUT are well-known PLS-complete problems. An instinctive approach to finding such a locally optimum solution is the FLIP method. Even though FLIP requires exponential time in worst-case instances, it tends to terminate quickly in practical instances. To explain this discrepancy, the run-time of FLIP has been studied in the smoothed complexity framework. Etscheid and Röglin (ACM Transactions on Algorithms, 2017) showed that the smoothed complexity of FLIP for MAX-CUT in arbitrary graphs is quasi-polynomial. Angel, Bubeck, Peres, and Wei (STOC, 2017) showed that the smoothed complexity of FLIP for MAX-CUT in complete graphs is  $O(\phi^5 n^{15.1})$ , where  $\phi$  is an upper bound on the random edge-weight density and  $n$  is the number of vertices in the input graph.

While Angel, Bubeck, Peres, and Wei's result showed the first polynomial smoothed complexity, they also conjectured that their run-time bound is far from optimal. In this work, we make substantial progress toward improving the run-time bound. We prove that the smoothed complexity of FLIP for MAX-CUT in complete graphs is  $O(\phi n^{7.83})$ . Our results are based on a carefully chosen matrix whose rank captures the run-time of the method along with improved rank bounds for this matrix and an improved union bound based on this matrix. In addition, our techniques provide a general framework for analyzing FLIP in the smoothed framework. We illustrate this general framework by showing that the smoothed complexity of FLIP for MAX-3-CUT in complete graphs is polynomial and for MAX- $k$ -CUT in arbitrary graphs is quasi-polynomial. We believe that our techniques should also be of interest toward showing smoothed polynomial complexity of FLIP for MAX- $k$ -CUT in complete graphs for larger constants  $k$ .

CCS Concepts: • **Theory of computation** → **Graph algorithms analysis; Randomness, geometry and discrete structures**; • **Mathematics of computing** → **Probability and statistics**;

Additional Key Words and Phrases: Smoothed analysis, local search, max-cut, k-cut

## ACM Reference format:

Ali Bibak, Charles Carlson, and Karthekeyan Chandrasekaran. 2021. Improving the Smoothed Complexity of FLIP for Max Cut Problems. *ACM Trans. Algorithms* 17, 3, Article 19 (July 2021), 38 pages.

<https://doi.org/10.1145/3454125>

An extended abstract of this work appeared in the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2019). This article contains complete proofs in addition to bounds on expected number of flips. Karthekeyan is supported in part by NSF CCF 18-14613.

Authors' addresses: A. Bibak and K. Chandrasekaran, University of Illinois, Urbana-Champaign, 104 S. Mathews Ave, Urbana, IL 61801; emails: {bibakse2, karthe}@illinois.edu; C. Carlson, University of Colorado Boulder, 1111 Engineering Drive, ECOT 717, 430 UCB Boulder, CO 80309; email: Charles.Carlson@Colorado.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2021 Association for Computing Machinery.

1549-6325/2021/07-ART19 \$15.00

<https://doi.org/10.1145/3454125>

## 1 INTRODUCTION

A  $k$ -cut in a graph is a partition of the vertex set into  $k$  parts. Given an edge-weighted graph and a  $k$ -cut, the value of the cut is the total weight of the edges crossing the partition. In the max- $k$ -cut problem, denoted MAX- $k$ -CUT, we are given a graph with edge weights and the goal is to find a  $k$ -cut with maximum value. For convenience, we will denote MAX-2-CUT as MAX-CUT. MAX-CUT is a well-known NP-hard problem whose study brought forth many new algorithmic techniques. In this work, we analyze the run-time of a local algorithm for MAX-CUT and more generally, for MAX- $k$ -CUT.

A  $k$ -cut is said to be a *local max- $k$ -cut* if the cut value cannot be improved by changing the part of any *single* vertex. We recall that a local max- $k$ -cut is, in fact, a  $(1 - 1/k)$ -approximate max- $k$ -cut [11]. The computation of a local max- $k$ -cut is of interest in game theory as it is a Nash-equilibrium in the party affiliation game [3, 8]: consider  $n$  players with certain weights between pairs of players and they would like to form  $k$  teams. The payoff for a player is the total weight of the edges between her and the players in her  $k - 1$  opposing teams. A local max- $k$ -cut is a Nash equilibrium for this game. Schäffer and Yannakakis showed that computing a local max-cut is likely to be intractable. In particular, they showed that it is **Polynomial-Time Local Search (PLS)**-complete [12].

A natural algorithm to find a local max- $k$ -cut is to start from an arbitrary  $k$ -cut and repeatedly perform local improving moves as long as possible. This is known as the FLIP method: it starts from an arbitrary  $k$ -cut and repeatedly increases the weight of the cut by moving a vertex from its current part to one of the other  $k - 1$  parts as long as such an improvement is possible. We note that an implementation of the FLIP method should specify how to choose (1) the initial  $k$ -cut, (2) the vertex to move in each iteration (if there is more than one vertex whose movement improves the cut value), and (3) the part to which the chosen vertex should be moved (if there is more than one choice that improves the cut value). The complexity of the FLIP method is the number of moves required for any implementation of the FLIP method to terminate.

The FLIP method corresponds to a natural player dynamics in the party affiliation game and hence its convergence time has been of much interest. Unfortunately, there are instances for which the FLIP method needs an exponential sequence of moves to converge even for MAX-CUT [2, 12]. Yet, empirical evidence suggests that the FLIP method is very fast in real-world instances of MAX-CUT [10].

The smoothed complexity framework, introduced by Spielman and Teng [13], is well-suited to explain the discrepancy in the performance of an algorithm between worst-case and practical instances. In the smoothed complexity framework for MAX- $k$ -CUT, we are given a graph  $G = (V, E)$  on  $n$  vertices along with a distribution  $f_e : [-1, 1] \rightarrow [0, \phi]$  according to which the weight  $X_e$  of edge  $e \in E$  is chosen. We note that the edge weights are independently distributed. Here, the parameter  $\phi$  determines the amount of random noise in the instance: if  $\phi \rightarrow \infty$ , then the instance is essentially a worst-case instance, whereas finite  $\phi$  amounts to some randomness in the instance. The restriction of the edge weights to the interval  $[-1, 1]$  is without loss of generality as arbitrary bounded weights can be scaled. The goal is to determine the run-time of the FLIP method in expectation (or with high probability) over the random choice of edge weights.

We distinguish between the two models under which the smoothed complexity of the FLIP method for MAX-CUT has been studied in the literature. In the case of smoothed complexity for *arbitrary graphs*, noise is added only to the existing edges of the given graph. In the case of smoothed complexity in *complete graphs*, noise is added to all vertex pairs including the non-edges of the given graph (by treating them as zero weight edges). We emphasize that this kind of subtlety between arbitrary graphs and complete graphs while studying the smoothed complexity has been prevalent—indeed, Spielman and Teng’s work analyzed the smoothed complexity of the simplex

method when noise is added to every entry of the constraint matrix including the zero entries; determining the smoothed complexity of the simplex method when noise is added only to the non-zero entries of the constraint matrix still remains as an important open problem.

All previous works [1, 5–7] have studied the complexity of the FLIP method only for MAX-CUT while its smoothed complexity for MAX- $k$ -CUT for  $k \geq 3$  has not been considered in the literature. We now mention the results relevant to this work. Etscheid and Röglin [7] showed that any implementation of the FLIP method for MAX-CUT in arbitrary graphs terminates using at most  $\phi n^{O(\log n)}$  moves with high probability, i.e., the smoothed complexity when a small amount of noise is added to every edge of the given graph is quasi-polynomial. Subsequently, Angel, Bubeck, Peres, and Wei [1] showed that any implementation of the FLIP method for MAX-CUT in complete graphs terminates using at most  $O(\phi^5 n^{15.1})$  moves with high probability, i.e., the smoothed complexity when a small amount of noise is added to every vertex pair is polynomial.

### 1.1 Our Results

Motivated by empirical evidence, Angel, Bubeck, Peres, and Wei conjectured that the dependence on  $n$  should be quasi-linear and therefore, raised the question of improving the run-time analysis. In this work, we address this question by improving the run-time analysis of the FLIP method for MAX-CUT.

**THEOREM 1.1.** *Let  $G = (V, E)$  be the complete graph on  $n$  vertices, and suppose that the edge weights  $(X_e)_{e \in E}$  are independent random variables chosen according to a probability density function  $f_e : [-1, 1] \rightarrow [0, \phi]$  for some  $\phi > 0$ . Then, for every constant  $\eta > 0$ ,*

- (i) *with high probability every implementation of the FLIP method for MAX-CUT terminates in at most  $23,400\phi n^{3+(2+\sqrt{2})(\sqrt{2}+\eta)} = O(\phi n^{7.829+3.414\eta})$  steps.*
- (ii) *the expected number of steps in every implementation of the FLIP method for MAX-CUT is  $O(\phi n^{\frac{47}{6}+\eta})$ .*

In particular, by taking  $\eta = 0.0001$ , our theorem implies that the FLIP method for MAX-CUT terminates in at most  $O(\phi n^{7.833})$  moves with high probability. This run-time bound improves the dependence on both the max-density  $\phi$  as well as the number of vertices  $n$  in comparison to that of Angel, Bubeck, Peres, and Wei [1]. The outline of our analysis technique follows the recipe introduced by Etscheid and Röglin [7] and followed by Angel, Bubeck, Peres, and Wei [1]: We associate a suitable matrix with the sequence of FLIP moves; we show a union bound that a long sequence of moves followed by the FLIP method will lead to a large improvement in the cut value with high probability provided that the rank of this associated matrix is large; next, we show that the rank of this matrix is indeed large. Our techniques differ from those of Angel, Bubeck, Peres, and Wei along three fronts: (1) the matrix that we associate with the sequence of moves is different from the one used by Angel, Bubeck, Peres, and Wei, but it is closer to the one used by Etscheid and Röglin, (2) the union bound that we show is much stronger than the one by Angel, Bubeck, Peres, and Wei, and (3) the rank lower bound that we show is much stronger than the one shown by Etscheid and Röglin.

Next, we turn to the smoothed complexity of the FLIP method for MAX- $k$ -CUT for  $k \geq 3$ . As mentioned above, all known results on the smoothed analysis of the FLIP method address only the case of  $k = 2$ . However, the convergence of this method for MAX- $k$ -CUT when  $k$  is larger is also of interest (for instance, from the perspective of understanding the natural dynamics in the party affiliation game). We take the first step towards this goal by analyzing the smoothed complexity of the FLIP method for MAX-3-CUT in complete graphs.

**THEOREM 1.2.** *Let  $G = (V, E)$  be the complete graph on  $n$  vertices, and suppose that the edge weights  $(X_e)_{e \in E}$  are independent random variables chosen according to a probability density function  $f_e : [-1, 1] \rightarrow [0, \phi]$  for some  $\phi > 0$ . Then,*

- (i) *for every constant  $\eta > 0$ , with high probability every implementation of the FLIP method for MAX-3-CUT terminates in at most  $O(\phi n^{99+\eta})$  steps.*
- (ii) *the expected number of steps in every implementation of the FLIP method for MAX-3-CUT is  $O(\phi n^{99})$ .*

In particular, by taking  $\eta = 0.1$ , our theorem implies that the FLIP method for MAX-3-CUT terminates in at most  $O(\phi n^{99.1})$  moves with high probability.<sup>1</sup> We observe that the techniques of Angel, Bubeck, Peres, and Wei do not extend to address the smoothed complexity of the FLIP method for MAX-3-CUT in complete graphs. Their union bound argument for the matrix that they associate with the sequence of moves crucially relies on the fact that its entries are independent of the starting 2-cut. Unfortunately, a similar matrix for MAX-3-CUT has entries that depend on the starting 3-cut and hence, their union bound arguments fail to extend. We overcome this issue by relying on a completely different matrix. We believe that our techniques underlying Theorem 1.2 provide a general framework to address the smoothed complexity of the FLIP method for MAX- $k$ -CUT in complete graphs for any constant  $k$ .

Finally, we also show that the smoothed complexity of the FLIP method for MAX- $k$ -CUT in arbitrary graphs (i.e., noise is added only to the edges of the given graph) is quasi-polynomial for constant  $k$ .

**THEOREM 1.3.** *Let  $G = (V, E)$  be an arbitrary graph on  $n$  vertices, and suppose that the edge weights  $(X_e)_{e \in E}$  are independent random variables chosen according to a probability density function  $f_e : [-1, 1] \rightarrow [0, \phi]$  for some  $\phi > 0$ . Then, for every constant  $\eta > 0$ ,*

- (i) *with high probability every implementation of the FLIP method for MAX- $k$ -CUT terminates in at most  $O(\phi n^{2(2k-1)k \log(kn)+3+\eta})$  steps.*
- (ii) *the expected number of steps in every implementation of the FLIP method for MAX- $k$ -CUT is  $O(\phi n^{2(2k-1)k \log(kn)+3+\eta})$ .*

## 1.2 Related Work

The literature of smoothed analysis is vast with successful analysis of several algorithms for various problems. We mention the works closely related to MAX-CUT. Elsässer and Tscheuschner [5] showed that if the edge weights are perturbed using Gaussian noise in graphs with maximum degree  $O(\log n)$ , then the complexity of the FLIP method is polynomial. Etscheid and Röglin [6] considered another special case in which the vertices are points in a  $d$ -dimensional space and the edge weights are given by the squared Euclidean distance between these points. In this setting, they showed that if the points are perturbed by Gaussian noise, then the complexity of the FLIP method is polynomial. After these special settings, Etscheid and Röglin [7] showed that the smoothed complexity of the FLIP method in arbitrary graphs is quasi-polynomial. Subsequently, Angel, Bubeck, Peres, and Wei [1] showed that the smoothed complexity of the FLIP method in complete graphs is polynomial.

As we mentioned in the introduction, the FLIP method is of interest as a natural dynamics towards computing a Nash equilibrium in certain games. In the *non-smoothed* setting, computing a

<sup>1</sup>Our run-time bound could be improved (see Remark 7), but we state a weaker bound for the purpose of simplicity in the analysis.

local MAX- $k$ -CUT is a special case of computing a pure Nash equilibrium in network coordination games. Independent of our work, Boodaghians, Kulkarni, and Mehta [4] have given an efficient algorithm for computing a Nash equilibrium in *smoothed* network coordination games. However, it is important to note that their smoothed setting for network coordination games when specialized to the case of MAX- $k$ -CUT *does not* correspond to our smoothed setting for MAX- $k$ -CUT. So, our results on the smoothed complexity of the FLIP method for local MAX-3-CUT and local MAX- $k$ -CUT *do not* follow from their results. Alternatively, they present a “smoothness preserving reduction” from computing Nash equilibrium in network coordination games involving only 2 strategies to computing local max-cut. Our Theorem 1.1 complements this result as it now follows that the smoothed complexity of a natural dynamics for computing a Nash equilibrium in 2-strategy network coordination games is  $O(n^{7.83})$ .

### 1.3 Preliminaries

All graphs considered in this work are simple. Let  $H = (V, E)$  be a directed graph and let  $v \in V$ . Then, we denote the outgoing neighborhood of  $v$  in  $H$  by  $\Delta_H^{out}(v) := \{u \in V \mid vu \in E\}$  and the incoming neighborhood of  $v$  in  $H$  by  $\Delta_H^{in}(v) := \{u \in V \mid uv \in E\}$ . Likewise, we denote the outgoing arcs from  $v$  in  $H$  by  $\delta_H^{out}(v) := \{wu \in E \mid w = v\}$  and the incoming arcs to  $v$  in  $H$  by  $\delta_H^{in}(v) := \{uw \in E \mid w = v\}$ . For a matrix  $M$ , we denote the element in the  $i$ th row and  $j$ th column of  $M$  by  $M[i, j]$ , and we denote the  $k$ th column of  $M$  by  $M^k$ . For two vectors  $a, b \in \mathbb{R}^n$ , we denote their dot product by  $\langle a, b \rangle$ . We denote the set of integers between 1 and an integer  $n$  by  $[n] := \{1, \dots, n\}$ .

## 2 OUTLINE OF OUR ANALYSIS

We will first describe our analysis for MAX-CUT followed by the analysis for MAX- $k$ -CUT.

### 2.1 Outline for MAX-CUT

In this section, we outline our proof of Theorem 1.1. Let  $G = (V, E)$  be the complete graph on  $n$  vertices and let  $X \in [-1, 1]^E$  be the edge weights. We will show that every sequence of improving moves of sufficiently large linear length, say  $2n$ , from any initial cut must increase the cut value by  $\Omega(\phi^{-1}n^{-4.83})$  with high probability. As the edge weights are bounded by at most 1, the value of every local max-cut is at most  $n^2$ . Hence, the FLIP method must terminate in at most  $O(\phi n^{7.83})$  moves with high probability. We now outline our proof showing that with high probability every linear length sequence of improving moves from any initial cut must increase the cut value by  $\Omega(\phi^{-1}n^{-4.83})$ .

We recall that a cut has two parts. So, we will represent a cut by a configuration in  $\{\pm 1\}^V$  (that represents the part of each vertex). A move of a vertex from a given configuration can be described using a vector  $\alpha \in \{0, \pm 1\}^E$  such that the increase in the cut value for that move is given by  $\langle \alpha, X \rangle$ . For a sequence of moves  $L$  from starting configuration  $\tau_0 \in \{\pm 1\}^V$ , we define a matrix  $P_{L, \tau_0}$  that conveniently nullifies the effect of non-moving vertices of  $L$ : for each pair of closest moves of a vertex  $v$  in  $L$ , we have a column in  $P_{L, \tau_0}$ , which is the sum of the two vectors  $\alpha_1$  and  $\alpha_2$  corresponding to those two moves of  $v$ . The main advantage of this matrix  $P_{L, \tau_0}$  is that it depends only on the starting configuration of the vertices that move in  $L$  and is independent of the starting configuration of the vertices that do not move in  $L$  (Proposition 1). This feature is crucially helpful while taking a union bound later. Our matrix  $P_{L, \tau_0}$  is also implicitly used by Etscheid and Röglin.

We say that a sequence  $L$  of moves is  $\epsilon$ -slow from an initial configuration  $\tau_0$  with respect to edge-weights  $X$  if each move of  $L$  strictly improves the cut value and moreover, the total improvement made by  $L$  to the cut-value is at most  $\epsilon$ . Thus, if  $L$  is  $\epsilon$ -slow from  $\tau_0$  with respect to  $X$ , then

$\langle \alpha_t, X \rangle > 0$  and  $\sum_{t=1}^{\text{length}(L)} \langle \alpha_t, X \rangle \leq \epsilon$ , where  $\alpha_t$  is the vector associated with the  $t$ th step of  $L$ . It follows that if  $L$  is  $\epsilon$ -slow from  $\tau_0$  with respect to  $X$ , then  $\langle C, X \rangle > 0$  for all columns  $C$  in  $P_{L, \tau_0}$  and moreover  $\sum_{C \in \text{Columns}(P_{L, \tau_0})} \langle C, X \rangle \leq 2\epsilon$ , since each column  $\alpha$  participates in at most two columns of the matrix  $P_{L, \tau_0}$ . We define the following event for a sequence  $L$  from a starting configuration  $\tau_0$  for edge weights  $X$ :  $\mathcal{E}_{L, \tau_0, X}$  is the event where  $\langle C, X \rangle > 0$  for all columns  $C$  in  $P_{L, \tau_0}$  and

$$\sum_{C \in \text{Columns}(P_{L, \tau_0})} \langle C, X \rangle \leq 2\epsilon.$$

We show that (Lemma 3.13) for a fixed choice of  $L$  and  $\tau_0$ , the probability (over the choices of  $X$ ) that  $\mathcal{E}_{L, \tau_0, X}$  happens is at most

$$\frac{(2\phi\epsilon)^{\text{rank}(P_{L, \tau_0})}}{\text{rank}(P_{L, \tau_0})!}.$$

Our probability bound mentioned above is stronger than the bound given by Etscheid and Röglin as well as Angel, Bubeck, Peres, and Wei. The main fact that we exploit to obtain this stronger bound is that an  $\epsilon$ -slow sequence, by definition, improves the total sum  $\sum_{C \in \text{Columns}(P_{L, \tau_0})} \langle C, X \rangle$  by at most  $2\epsilon$ . In contrast, previous works used  $\epsilon$ -slow in a weaker manner: Etscheid and Röglin only used the fact that  $\langle C, X \rangle \leq 2\epsilon$  for every column  $C$  of the matrix  $P_{L, \tau_0}$  while Angel, Bubeck, Peres, and Wei only used the fact that  $\langle \alpha_t, X \rangle \leq \epsilon$  for every step  $t$  in  $L$ . We deviate from their analysis to further exploit the definition of  $\epsilon$ -slowness.

Next, we need to upper bound the probability that there exists a starting configuration  $\tau_0$  and a linear length sequence  $L$  such that the event  $\mathcal{E}_{L, \tau_0, X}$  happens. To attempt the natural union bound, we need a lower bound on the rank of the matrix  $P_{L, \tau_0}$  for sequences of linear length. For example, if the rank is at least some constant fraction of  $n$ , then we may use a straightforward union bound. However, there are linear length sequences that have rank much smaller than  $n$ . We address this issue by focusing on *critical* sequences.

A sequence  $B$  is a critical sequence if  $\ell(B) = \lceil 1.71s(B) \rceil$  and moreover,  $\ell(B') < 1.71s(B')$  for every subsequence  $B'$  of  $B$ , where  $\ell(L)$  and  $s(L)$  denote the length and the number of vertices, respectively, in a sequence  $L$ . Here, the constant 1.71 is chosen optimally to obtain the best possible run-time using our technique and the reasoning behind this choice is not insightful for the purposes of this overview. Critical sequences were introduced by Angel, Bubeck, Peres, and Wei, who also showed that every sequence of length  $\lceil 1.71n \rceil$  (i.e., sufficiently large linear length) contains a critical subsequence (Lemma 3.14). Thus, it suffices to upper bound the probability that there exists a critical sequence  $B$  and a starting configuration  $\tau_0$  such that the event  $\mathcal{E}_{B, \tau_0, X}$  happens.

Let us now fix a critical sequence  $B$  and obtain an upper bound on the probability that there exists a starting configuration  $\tau_0$  such that the event  $\mathcal{E}_{B, \tau_0, X}$  happens. We exploit the fact that the event  $\mathcal{E}_{B, \tau_0, X}$  is independent of the starting configuration of the vertices that do not move in  $B$  (as the matrix  $P_{B, \tau_0}$  has this property). Thus, it suffices to perform a union bound over the starting configuration of the vertices that move in  $B$ . The number of such vertices is  $s(B)$ , hence the number of possible starting configurations of moving vertices is  $2^{s(B)}$ . Now, it remains to bound the probability that the event  $\mathcal{E}_{B, \tau_0, X}$  happens, where  $\tau_0$  is a fixed choice of the starting configuration of the vertices that move in  $B$  and an arbitrarily chosen starting configuration of the vertices that *do not* move in  $B$ . For this, the above discussions suggest that we need a lower bound on the rank of the matrix  $P_{B, \tau_0}$  for a critical sequence  $B$ . We show that for a critical sequence  $B$ , the rank of  $P_{B, \tau_0}$  is at least  $0.38s(B)$  (Corollary 3.11). Thus, for a fixed critical sequence  $B$ , the probability that there exists a starting configuration  $\tau_0$  such that the event  $\mathcal{E}_{B, \tau_0, X}$  happens

is at most

$$2^{s(B)} \frac{(2\phi\epsilon)^{0.38s(B)}}{(0.38s(B))!}.$$

Finally, we take a union bound over critical sequences. The number of possible critical sequences  $B$  with  $s(B) = s$  is at most  $\binom{n}{s} s^{\lceil 1.71s \rceil}$ , since the length of such a critical sequence is  $\lceil 1.71s \rceil$ . Thus, the probability that there exists a critical sequence  $B$  and a starting configuration  $\tau_0$  such that the event  $\mathcal{E}_{B, \tau_0, X}$  happens is at most

$$\sum_{s=1}^n \binom{n}{s} s^{\lceil 1.71s \rceil} \left( 2^s \frac{(2\phi\epsilon)^{0.38s}}{(0.38s)!} \right) = o(1),$$

when  $\epsilon = \phi^{-1} n^{-4.83}$  using Stirling's approximation.

Our analysis approach builds on top of two previously known ingredients: the matrix  $P_{L, \tau_0}$  implicitly used by Etscheid and Röglin [7] to nullify the effect of non-moving vertices and the notion of a critical block introduced by Angel, Bubeck, Peres, and Wei [1] that is helpful to show a lower bound on the rank of the relevant matrix. Our main contributions to improve the run-time analysis are (1) a tighter union bound by exploiting the full power of  $\epsilon$ -slowness and (2) improved rank lower bounds for critical sequences.

## 2.2 Outline for MAX-3-CUT and MAX- $k$ -CUT in Arbitrary Graphs

In this section, we outline our proof of Theorems 1.2 and 1.3. The high-level approach is similar to the one for MAX-CUT described above. Let  $G = (V, E)$  be an *arbitrary* graph on  $n$  vertices and let  $X \in [-1, 1]^E$  be the edge weights. Our goal is to show that every sufficiently long linear length sequence of improving moves from any initial  $k$ -cut must increase the cut value by some non-negligible amount.

We represent a  $k$ -cut by a configuration in  $[k]^V$  that represents the part of each vertex. Using an observation by Frieze and Jerrum [9], we can again associate a vector  $\alpha \in \{0, \pm 1\}^E$  with each move such that the increase in the cut value for that move is given by  $\langle \alpha, X \rangle$ . We emphasize that these vectors depend on the starting configuration. The analysis technique by Angel, Bubeck, Peres, and Wei for MAX-CUT does not extend to MAX- $k$ -CUT even for  $k = 3$  primarily due to the dependence of the rank of the matrix of these vectors on the starting configuration.

Our main tool to overcome this issue is by considering an appropriate matrix  $P_{L, \tau_0}$  for MAX- $k$ -CUT. Instead of pairs of nearest moves in  $L$  that was helpful for MAX-CUT, here, we define the notion of a *cycle* over a vertex in  $L$ . A cycle over a vertex is a set of moves of that vertex that result in the vertex moving from one part and eventually returning to that same part. We note that this generalizes the concept of a pair in the MAX-CUT setting, since any two nearest moves would form a cycle. Next, we define the matrix  $P_{L, \tau_0}$  to have a column for every cycle in  $L$  that is the sum of all the vectors that correspond to moves in that cycle. The matrix  $P_{L, \tau_0}$  has the property that it is independent of the starting configuration of the vertices that do not move in  $L$  (Proposition 2).

We use the same notion of  $\epsilon$ -slowness as that for MAX-CUT. Now, if  $L$  is  $\epsilon$ -slow from an initial configuration  $\tau_0$  with respect to edge weights  $X$ , then  $\langle C, X \rangle \in (0, k\epsilon]$  for all columns  $C \in P_{L, \tau_0}$ , since each column of the matrix  $P_{L, \tau_0}$  is the sum of at most  $k$  vectors  $\alpha_{t_1}, \dots, \alpha_{t_k}$ . We define the following event for a sequence  $L$  from a starting configuration  $\tau_0$  for edge weights  $X$ :

$$\mathcal{D}_{L, \tau_0, X} : \langle C, X \rangle \in (0, k\epsilon] \text{ for all columns } C \text{ in } P_{L, \tau_0}.$$

For a fixed  $L$  and  $\tau_0$  (Lemma 4.16), the probability (over the choices of  $X$ ) that  $\mathcal{D}_{L, \tau_0, X}$  happens is at most

$$(k\phi\epsilon)^{\text{rank}(P_{L, \tau_0})}.$$

The above result follows from Lemma A.1 in Reference [7].

Next, we need a lower bound on the rank of the matrix  $P_{L, \tau_0}$ . We emphasize that this needs substantially new combinatorial ideas in the form of considering the many different ways a cycle can interact with other cycles and vertices that are not part of any cycle.

- For the complete graph in the MAX-3-CUT case, we show that the rank of a critical improving sequence  $B$  is at least  $(1/32)s(B)$ , where  $s(B)$  is the number of vertices that move in the sequence  $B$  (Corollary 4.14). With this rank lower bound and the probability bound mentioned above, the rest of the analysis for the complete graph is similar to the analysis for MAX-CUT.
- For arbitrary graphs in the MAX- $k$ -CUT case, we show that the rank of an improving sequence  $B$  is at least half the number of vertices that appear in a cycle of  $B$  (Lemma 4.7). Next, we show that a sequence  $B$  of length  $kn$  must have a subsequence  $B'$  with at least  $1/(2(2k-1)k \log(kn))s(B)$  vertices that appear in some cycle of  $B'$ . We use these results in conjunction with the probability bound mentioned above to show that every sequence of improving moves of length  $kn$  from any initial  $k$ -cut must increase the cut value by  $\Omega(\phi^{-1}n^{-2(2k-1)k \log kn})$ .

Our main contributions are twofold: (1) We introduce the appropriate matrix  $P_{L, \tau_0}$  that nullifies the effect of non-moving vertices, which is crucial to perform the union bound, and (2) we show rank lower bounds for this matrix for  $k$ -cut in arbitrary graphs and for 3-cut in the complete graph based on combinatorial arguments. We believe that our techniques should also be helpful to show polynomial smoothed complexity of FLIP for MAX- $k$ -CUT for larger constants  $k$  in the complete graph case.

### 3 SMOOTHED ANALYSIS OF FLIP FOR MAX-CUT

In this section, we prove Theorem 1.1. We begin with some notation.

Let  $G = (V, E)$  be a connected graph with  $n$  vertices and let  $X : E \rightarrow [-1, 1]$  be an edge-weight function. We recall a convenient formulation of the objective function for MAX-CUT. We consider the space  $\{\pm 1\}^V$  of configurations that define a partition of the vertex set into two parts. For a configuration  $\tau \in \{\pm 1\}^V$ , we denote the part of  $v$  by  $\tau(v)$ . For a configuration  $\tau \in \{\pm 1\}^V$ , the weight of  $\tau$  is given by

$$\frac{1}{2} \sum_{uv \in E} X_{uv} (1 - \tau(u)\tau(v)). \quad (1)$$

Let

$$H(\tau) := -\frac{1}{2} \sum_{uv \in E} X_{uv} \tau(u)\tau(v).$$

We observe that  $H(\tau)$  is a translation of (1) by half the total weight of all edges and hence, it suffices to work with  $H(\tau)$  henceforth.

We analyze the run-time of the FLIP method in the smoothed framework for the complete graph. We will work with the complete graph in this section and avoid stating this explicitly henceforth. A *flip* of a vertex  $v \in V$  changes  $\tau(v)$  to  $-\tau(v)$ . We will denote a *move* by the vertex that is flipped. We will need the notion of improving sequences that we define now.

*Definition 3.1.* Let  $L$  be a sequence of moves,  $\tau_0 \in \{\pm 1\}^V$  be an initial configuration and  $X \in [-1, 1]^E$  be the edge weights. We will denote the length of the sequence  $L$  by  $\ell(L)$ , the set of vertices appearing in the moves in  $L$  by  $S(L)$ , and  $s(L) := |S(L)|$ . We will denote the  $t$ th move of  $L$  by  $L(t) = v_t$ .

- (1) For each  $t \in [\ell(L)]$ , we will denote  $\tau_t$  as the configuration obtained from  $\tau_{t-1}$  by setting  $\tau_t(u) := \tau_{t-1}(u)$  for every  $u \in V \setminus \{v_t\}$  and  $\tau_t(v_t) := -\tau_{t-1}(v_t)$ .

- (2) We say that  $L$  is *improving from  $\tau_0$  with respect to  $X$*  if  $H(\tau_t) - H(\tau_{t-1}) > 0$  for all  $t \in [\ell(L)]$ . We say that  $L$  is  *$\epsilon$ -slowly improving from  $\tau_0$  with respect to  $X$*  if  $L$  is improving from  $\tau_0$  and  $H(\tau_{\ell(L)}) - H(\tau_0) \in (0, \epsilon]$ .

The improvement of  $H(\tau)$  in a step can be written as the inner product between the edge-weight vector  $X$  and a suitable  $\{0, \pm 1\}$ -vector. We formalize this with a convenient matrix below.

*Definition 3.2.* Let  $L$  be a sequence of moves and  $\tau_0 \in \{\pm 1\}^V$  be an initial configuration. Let  $M_{L, \tau_0} \in \{0, \pm 1\}^{E \times [\ell(L)]}$  be a matrix with rows corresponding to the edges of  $G$ , columns corresponding to time-steps in the sequence  $L$ , and whose entries are given by

$$M_{L, \tau_0}[\{a, b\}, t] := \begin{cases} +1 & \text{if } L(t) \in \{a, b\} \text{ and } \tau_t(a) \neq \tau_t(b), \\ -1 & \text{if } L(t) \in \{a, b\} \text{ and } \tau_t(a) = \tau_t(b), \\ 0 & \text{otherwise,} \end{cases}$$

where  $\{a, b\} \in E$  and  $t \in [\ell(L)]$ .

*Remark 1.* For a sequence  $L$  from an initial configuration  $\tau_0$ , we have  $H(\tau_t) - H(\tau_{t-1}) = \langle M_{L, \tau_0}^t, X \rangle$ .

Next, we need the notion of repeating and singleton vertices. We note that the following definition does not depend on the initial configuration.

*Definition 3.3.* Let  $L$  be a sequence of moves. We will denote the number of times that a vertex  $v$  moves in  $L$  by  $\#_L(v)$ . A vertex  $v \in S(L)$  is called *repeating* if  $\#_L(v) \geq 2$  and is called a *singleton* otherwise. Let  $S_1(L)$  and  $S_2(L)$  denote the set of singleton and repeating vertices of  $L$ , respectively, and let  $s_1(L) := |S_1(L)|$  and  $s_2(L) := |S_2(L)|$ . For time steps  $t_1, t_2 \in [\ell(L)]$ , the ordered pair  $(t_1, t_2)$  is a *pair for vertex  $v \in V$*  if  $t_1 < t_2$ ,  $L(t_1) = v = L(t_2)$  and  $L(t) \neq v$  for all  $t \in \{t_1 + 1, \dots, t_2 - 1\}$ . For all  $v \in V$ , let  $\Gamma(L, v)$  be the set of pairs for  $v$  in  $L$  and  $\Gamma(L) := \cup_{v \in V} \Gamma(L, v)$  be the set of pairs for all vertices in  $L$ .

We now define a suitable matrix that will nullify the influence of non-moving vertices.

*Definition 3.4.* Let  $L$  be a sequence of moves and let  $\tau_0 \in \{\pm 1\}^V$  be an initial configuration. Let  $P_{L, \tau_0} \in \{0, \pm 1\}^{E \times \Gamma(L)}$  be a matrix with rows corresponding to edges of  $G$ , columns corresponding to pairs in  $L$ , and whose entries are given by

$$P_{L, \tau_0}[\{a, b\}, C] := \sum_{t \in C} M_{L, \tau_0}[\{a, b\}, t],$$

where  $\{a, b\} \in E$  and  $C \in \Gamma(L)$ .

**PROPOSITION 1.** *Let  $L$  be a sequence of moves. If  $v \in V \setminus S(L)$ , then  $P_{L, \tau_0}[\{a, v\}, C] = 0$  for every  $C \in \Gamma(L)$  and  $\{a, v\} \in E$ .*

**PROOF.** Let  $C = (t_1, t_2) \in \Gamma(L)$  and  $\{a, v\} \in E$ . Since  $v$  is not in  $S(L)$ , it follows that  $C$  is not a pair for  $v$ . If  $C$  is not a pair for  $a$ , then  $M_{L, \tau_0}[\{a, v\}, t_i] = 0$  for  $i \in [2]$  and hence  $P_{L, \tau_0}[\{a, v\}, C] = 0$ . Suppose  $C$  is a pair for  $a$ . Then, it follows that  $M_{L, \tau_0}[\{a, v\}, t_1] = -M_{L, \tau_0}[\{a, v\}, t_2]$  and hence  $P_{L, \tau_0}[\{a, v\}, C] = M_{L, \tau_0}[\{a, v\}, t_1] + M_{L, \tau_0}[\{a, v\}, t_2] = 0$ .  $\square$

### 3.1 Rank Lower Bounds for $P_{L, \tau_0}$

Let  $L$  be a sequence of moves and let  $\tau_0$  be an initial configuration. In this section, we show a lower bound on the rank of  $P_{L, \tau_0}$ . For this, we will relate the rank of the matrix  $P_{L, \tau_0}$  with the number of edges of a certain directed graph. This graph has certain properties that imply this relation.

These properties, in fact, motivate the definition of the graph; hence, we define the properties in Definition 3.5 below.

First, we need non-zero entries in  $P_{L, \tau_0}$  to obtain rank lower bounds. For a pair  $C = (t_1, t_2) \in \Gamma(L)$  for vertex  $v$ , the entry  $P_{L, \tau_0}[\{u, v\}, C]$  is non-zero if and only if  $u$  appears an odd number of times between  $t_1$  and  $t_2$ . This motivates the definition of  $L$ -good-arcs below. Second, we need several linearly-independent rows in the matrix  $P_{L, \tau_0}$ . The notion of  $L$ -neighbor-wise independence below will help us to obtain an upper-triangular submatrix of  $P_{L, \tau_0}$ , thereby allowing us to get linearly-independent rows. Finally, the notion of *functional* graphs below will help us later in constructing an  $L$ -good  $L$ -neighbor-wise independent directed acyclic graph with a large number of edges. We define these notions below.

*Definition 3.5.* Let  $L$  be a sequence of moves.

- (i) For  $u, v \in S(L)$ , we will call the ordered pair  $vu$  to be an  $L$ -good-arc if there exists a pair  $C \in \Gamma(L)$  for  $v$  such that  $P_{L, \tau_0}[\{u, v\}, C] \neq 0$ . A directed graph  $H$  with node set  $S(L)$  is  $L$ -good if every arc in  $H$  is an  $L$ -good-arc.
- (ii) A directed graph  $H$  with node set  $S(L)$  is  $L$ -neighbor-wise independent if for every  $v \in S(L)$ , there exists an ordering of  $\Delta_H^{out}(v)$ , say  $u_1, \dots, u_m$ , along with pairs  $C_1, \dots, C_m \in \Gamma(L)$  for  $v$  such that for every  $i \in [m]$ ,
  - (a)  $P_{L, \tau_0}[\{v, u_i\}, C_i] \neq 0$  and
  - (b)  $P_{L, \tau_0}[\{v, u_j\}, C_i] = 0$  for every  $j \in \{i + 1, \dots, m\}$ .
- (iii) A directed graph  $H$  with node set  $S(L)$  is *functional* if  $|\delta_H^{out}(v)| = 1$  for every  $v \in S_2(L)$ .

*Remark 2.* For  $u, v \in S(L)$ , the ordered pair  $vu$  is an  $L$ -good-arc if and only if there exists a pair  $(t_1, t_2) \in \Gamma(L)$  for  $v$  such that the number of times  $u$  appears between  $t_1$  and  $t_2$  is odd.

The following lemma is our key tool in obtaining a lower bound on the rank of  $P_{L, \tau_0}$ . We show that the rank is at least the number of arcs in an  $L$ -good  $L$ -neighbor-wise independent directed acyclic graph.

**LEMMA 3.6.** *Let  $L$  be a sequence of moves and let  $\tau_0 \in \{\pm 1\}^V$  be an initial configuration. Let  $H$  be an  $L$ -good  $L$ -neighbor-wise independent directed acyclic graph. Then,*

$$\text{rank}(P_{L, \tau_0}) \geq |E(H)|.$$

**PROOF.** Consider the submatrix  $B_H$  of  $P_{L, \tau_0}$  consisting of the rows corresponding to edges  $\{u, v\}$  for every arc  $vu \in E(H)$ . We now show that the matrix  $B_H$  has full row-rank by induction on  $|E(H)|$ . The base case for  $|E(H)| = 0$  is trivial.

For the induction step, we consider  $|E(H)| \geq 1$ . Suppose that there exist coefficients  $\mu_{\{u, v\}} \in \mathbb{R}$  for every  $uv \in E(H)$  such that

$$\sum_{uv \in E(H)} \mu_{\{u, v\}} P_{L, \tau_0}[\{u, v\}, (t, t')] = 0$$

for every pair  $(t, t') \in \Gamma(L)$ . Since  $H$  is a directed acyclic graph with at least one arc, there exists a node  $v \in V(H)$  with  $|\delta_H^{out}(v)| \geq 1$  and  $|\delta_H^{in}(v)| = 0$ .

*Claim 1.* For every  $u \in \Delta_H^{out}(v)$ , the coefficient  $\mu_{\{v, u\}}$  is zero.

**PROOF.** Consider the ordering  $u_1, \dots, u_m$  of the vertices in  $\Delta_H^{out}(v)$  and pairs  $C_i \in \Gamma(L)$  satisfying Definition 3.5 (ii). We will show that  $\mu_{\{v, u_i\}} = 0$  for every  $i \in [m]$ . We use induction on  $j \in [m + 1]$  to show that  $\mu_{\{v, u_i\}} = 0$  for every  $i < j$ .

The base case for  $j = 1$  is trivial. To prove the inductive step, we consider  $j = j_0 + 1 \geq 2$ . By the induction hypothesis, we know that  $\mu_{\{v, u_i\}} = 0$  for every  $i < j_0$ . Thus,

$$0 = \sum_{i=j_0}^m \mu_{\{v, u_i\}} P_{L, \tau_0}[\{v, u_i\}, C_{j_0}].$$

Moreover, by the choice of  $C_{j_0}$ , we also have that

$$P_{L, \tau_0}[\{v, u_{j_0}\}, C_{j_0}] \neq 0, \text{ and } P_{L, \tau_0}[\{v, u_i\}, C_{j_0}] = 0 \forall i \in \{j_0 + 1, \dots, m\}.$$

Consequently, we obtain that  $\mu_{\{v, u_{j_0}\}} = 0$ . We have thus concluded the inductive statement that  $\mu_{\{v, u_i\}} = 0$  for every  $i < j_0 + 1$ .  $\square$

As a consequence of the claim, we have that the matrix  $B_H$  has full row-rank if and only if the matrix  $B_{H'}$  obtained from the graph  $H' := H - \delta_H^{out}(v)$  has full row-rank. We note that the graph  $H'$  is also an  $L$ -good  $L$ -neighbor-wise independent directed acyclic graph. Moreover,  $|E(H')| < |E(H)|$ . Thus, by the induction hypothesis, the matrix  $B_{H'}$  has full row-rank. Hence, the matrix  $B_H$  also has full row-rank.  $\square$

To show a lower bound on the rank of  $P_{L, \tau_0}$ , we switch our attention from  $L$  to specific *blocks* of  $L$  with special properties. We will see later that all sufficiently long  $L$  must contain such blocks.

*Definition 3.7.* Let  $L$  be a sequence of moves. A *block* of  $L$  is a continuous subsequence of  $L$ . For  $t_1, t_2 \in [\ell(L)]$ , we will denote the block of  $L$  between time-steps  $t_1$  and  $t_2$  as  $L[t_1, t_2]$ . For a block  $B$  of  $L$ , we will denote the starting and ending time-step of  $B$  in  $L$  as  $t_{beg}(B)$  and  $t_{end}(B)$ . Let  $\beta > 0$  be a constant. A block  $B$  in  $L$  is  $\beta$ -critical if  $\ell(B) \geq (1 + \beta)s(B)$  and  $\ell(B') < (1 + \beta)s(B')$  for every block  $B'$  that is strictly contained in  $B$ .

We begin with a simple property of  $\beta$ -critical blocks.

*Claim 2.* Let  $B$  be a  $\beta$ -critical block for  $0 < \beta \leq 1$ . Then, for every block  $B'$  that is strictly contained in  $B$ , we have that  $S_1(B') \neq \emptyset$ .

**PROOF.** Let  $B'$  be a block that is strictly contained in  $B$ . Suppose for the sake of contradiction that every vertex in  $B'$  appears at least twice. Then,  $\ell(B') \geq 2s(B') \geq (1 + \beta)s(B')$ , which contradicts the criticality of  $B$ .  $\square$

We use Claim 2 to show that for a  $\beta$ -critical block  $B$ , there exists a path of  $B$ -good-arcs from any repeating vertex to a singleton vertex.

*LEMMA 3.8.* Let  $B$  be a  $\beta$ -critical block for  $0 < \beta \leq 1$  with  $S_1(B) \neq \emptyset$ . Then, for every  $v \in S_2(B)$ , there exists a sequence  $u_0 u_1, u_1 u_2, \dots, u_k u_{k+1}$  of  $L$ -good-arcs such that  $u_0 = v$  and  $u_{k+1} \in S_1(B)$ .

**PROOF.** Let  $v \in S_2(B)$ . We will build a chain of blocks  $B_0 \subsetneq B_1 \subsetneq B_2 \subsetneq \dots \subsetneq B_k \subsetneq B$  and vertices  $u_0, u_1, \dots, u_k, u_{k+1}$  such that

- (i)  $u_0 = v$  and the vertices  $u_0, u_1, \dots, u_{k+1}$  are distinct,
- (ii)  $u_{k+1} \in S_1(B)$ ,
- (iii)  $u_i \in S_1(B_{i-1})$  for every  $i \in [k + 1]$ ,
- (iv) the start vertex and end vertex of the sequence  $B_i$  are  $u_i$  and a vertex in  $\{u_0, u_1, \dots, u_{i-1}\}$ ,  
and
- (v) for every  $i \in [k + 1]$ , there exists an  $L$ -good-arc  $u_j u_i$  for some  $j < i$ .

---

**Procedure**


---

**Input:**  $\beta$ -critical block  $B$  ( $0 < \beta \leq 1$ ) with  $S_1(B) \neq \emptyset$  and a vertex  $v \in S_2(B)$ .

- (1) Initialize  $i \leftarrow 0$ ,  $u_0 \leftarrow v$ ,  $B_0 \leftarrow B[t_1, t_2]$  for an arbitrary pair  $(t_1, t_2)$  for  $u_0$ .
- (2)  $u \leftarrow$  an arbitrary vertex from  $S_1(B_0)$  and  $r_u \leftarrow$  time-step  $r \in \{t_{beg}(B_0), \dots, t_{end}(B_0)\}$  for which  $B(r) = u$ .
- (3) While ( $u \notin S_1(B)$ ):
  - (a)  $i \leftarrow i + 1$ .
  - (b)  $u_i \leftarrow u$ .
  - (c) Let  $(r_u, q_u)$  (or  $(q_u, r_u)$ ) be a pair for  $u$ . Since  $u \notin S_1(B)$ , but  $u \in S_1(B_{i-1})$  it follows that  $q_u \notin [t_{beg}(B_{i-1}), t_{end}(B_{i-1})]$ . Set

$$B'_{i-1} \leftarrow \begin{cases} B[q_u, t_{beg}(B_{i-1}) - 1] & \text{if } q_u < t_{beg}(B_{i-1}), \\ B[t_{end}(B_{i-1}) + 1, q_u] & \text{if } q_u > t_{end}(B_{i-1}). \end{cases}$$

- (d)  $B_i \leftarrow$  concatenation of  $B_{i-1}$  and  $B'_{i-1}$  as they appear in  $B$ .
  - (e)  $u \leftarrow$  an arbitrary vertex from  $S_1(B_i)$  and  $r_u \leftarrow$  time-step  $r \in \{t_{beg}(B_i), \dots, t_{end}(B_i)\}$  for which  $B(r) = u$ .
  - (4)  $u_{i+1} \leftarrow u$ .
- 

Fig. 1. Procedure for Lemma 3.8.

We observe that such a chain immediately implies the existence of the required sequence of  $L$ -good-arcs that proves the lemma: by condition (v), there exists a sequence of  $L$ -good-arcs that form a path that ends at  $u_{k+1}$  and necessarily starts at  $u_0$ . By condition (i), we have that  $u_0 = v$  and by condition (ii), we have that  $u_{k+1} \in S_1(B)$ .

We now show that a chain of blocks and vertices satisfying properties (i)–(v) indeed exists. For this, we consider the procedure given in Figure 1. We first note that a valid choice for the vertex  $u$  in steps 2 and 3(e) exists by Claim 2 and by the fact that  $S_1(B) \neq \emptyset$ . The sequence constructed by the procedure is a nested chain, since  $B_{i+1}$  is a concatenation of  $B_i$  with a block adjacent to  $B_i$ . Moreover, by steps 3(c) and 3(d), the block  $B_{i-1}$  is strictly contained in the block  $B_i$ . Thus, the procedure will indeed terminate, since the sequence of blocks  $B_0, B_1, \dots, B_k$  forms a nested chain that grows in size. In the worst case, the procedure will terminate with  $B_k = B$  at which point step 3(e) finds  $u \in S_1(B_k) = S_1(B)$ . We now show that the required conditions are satisfied. Let the procedure terminate with the nested chain  $B_0 \subsetneq B_1 \subsetneq B_2 \subsetneq \dots \subsetneq B_k \subseteq B$  and vertices  $u_0, u_1, \dots, u_k, u_{k+1}$ .

- (i) By initialization  $u_0 = v$ . Since  $B_i$  is strictly contained in  $B_{i+1}$  and by steps 3(e) and 3(b), the vertices  $u_0, u_1, \dots, u_{k+1}$  are distinct.
- (ii) By the termination criteria, we have that  $u_{k+1} \in S_1(B)$ .
- (iii) By steps 3(e), 3(a) and 3(b), we have that  $u_i \in S_1(B_{i-1})$  for every  $i \in [k + 1]$ .
- (iv) By steps 3(c) and 3(d), the start vertex and end vertex of the sequence  $B_i$  are the vertex  $u_i$  and a vertex in  $\{u_0, u_1, \dots, u_{i-1}\}$ .
- (v) Let  $i \in [k + 1]$  and consider  $u_i$ . Let  $h$  be the smallest index in  $\{0, 1, \dots, k + 1\}$  such that  $u_i \in S_1(B_h)$ . Then, by condition (iii), we have that  $h \leq i - 1$ . We will show that the arc  $u_h u_i$  is an  $L$ -good-arc. If  $h = 0$ , then, by the choice of  $B_0$ , we have that the arc  $u_0 u_i$  is an  $L$ -good-arc. So, we may assume that  $h \geq 1$ . Then, the block  $B_h$  is a concatenation of  $B_{h-1}$  and  $B'_{h-1}$ . Also, by the choice of  $h$ , we know that  $u_i \notin B_{h-1}$ . Hence, the vertex  $u_i \in B'_{h-1}$ . Now, by condition (iv), we have that the block  $B'_{h-1}$  ends with  $u_h$ . By condition (iii), we have that  $u_h \in S_1(B_{h-1})$ . Hence, we have an  $L$ -good-arc  $u_h u_i$ .  $\square$

We need the notion of transition and singleton blocks.

*Definition 3.9.* Let  $L$  be an improving sequence from some initial configuration. Then a maximal block of repeating vertices is called a *transition block* and a maximal block of singletons is called a *singleton block*.

For a sequence  $L$  and a vertex  $v \in S_2(L)$ , let  $b_L(v)$  denote the number of transition blocks that contain  $v$ . Also, let  $R(L) := \{v \in S_2(L) : b_L(v) \geq 2\}$  (i.e.,  $R(L)$  is the set of vertices that appear in at least two transition blocks) and let  $r(L) := |R(L)|$ .

LEMMA 3.10. *Let  $B$  be a  $\beta$ -critical block where  $0 < \beta \leq 1$  with  $S_1(B) \neq \emptyset$ . Let  $\tau_0 \in \{\pm 1\}^V$  be an initial configuration. Then,*

$$\text{rank}(P_{B, \tau_0}) \geq s_2(B) - r(B) + \sum_{v \in S_2(B)} (b_B(v) - 1).$$

PROOF. We first build a functional  $B$ -good directed acyclic graph with  $s_2(B)$  arcs as follows: consider the  $B$ -good directed graph  $H$  over the node set  $S(B)$  containing all possible  $B$ -good-arcs. Now, run a reverse breadth first search from the nodes in  $S_1(B)$  and consider the subgraph  $H'$  over the node set  $S(B)$  obtained by including only the reverse-BFS-tree arcs. By Lemma 3.8, the reverse BFS search traverses all nodes in  $S_2(B)$  and hence, the graph  $H'$  has at least  $s_2(B)$  arcs. Moreover, the graph  $H'$  is a functional graph and acyclic, since we included only the reverse-BFS-tree arcs. Now, we note that every functional  $B$ -good directed graph is  $B$ -neighbor-wise independent as every node has only one outgoing arc. Thus, the graph  $H'$  is a  $B$ -good  $B$ -neighbor-wise independent directed acyclic graph.

Next, we add more arcs to  $H'$  to obtain a larger  $B$ -good  $B$ -neighbor-wise independent directed acyclic graph  $H''$  as follows: Consider a node  $v \in R(B)$ . We have exactly one outgoing arc  $vw$  from  $v$  in  $H'$ . For every two adjacent transition blocks containing  $v$ , we may add one  $B$ -good-arc  $vu$  to a vertex  $u$  that appears in a singleton block between those two adjacent transition blocks. If  $vw$  is not  $B$ -neighborwise-independent with the added arcs, then we discard the arc  $vw$ . We perform the above operation for every node  $v \in R(B)$ . The resulting graph  $H''$  is a  $B$ -good  $B$ -neighbor-wise independent directed acyclic graph. The number of arcs in  $H''$  is at least

$$s_2(B) + \sum_{v \in R(B)} (b_B(v) - 2) = s_2(B) - r(B) + \sum_{v \in S_2(B)} (b_B(v) - 1).$$

Then, by Lemma 3.6, we have that

$$\text{rank}(P_{B, \tau_0}) \geq s_2(B) - r(B) + \sum_{v \in S_2(B)} (b_B(v) - 1). \quad \square$$

COROLLARY 3.11. *Let  $B$  be a  $\beta$ -critical block where  $0 < \beta \leq 1$  with  $S_1(B) \neq \emptyset$ . Let  $\tau_0 \in \{\pm 1\}^V$  be an initial configuration. Then,*

$$\text{rank}(P_{B, \tau_0}) \geq \frac{\beta}{1 + 2\beta} s(B).$$

PROOF. We first show that  $\text{rank}(P_{B, \tau_0}) \geq \frac{\beta}{1+\beta} s_1(B)$ . Let  $T_1, \dots, T_k$  denote the transition blocks of  $B$ , and for a vertex  $v$ , let  $1_{v \in T_i}$  denote the indicator function for whether the vertex  $v$  appears in  $T_i$ . Then,

$$\sum_{v \in S_2(B)} b_B(v) = \sum_{v \in S_2(B)} \sum_{i=1}^k 1_{v \in T_i} = \sum_{i=1}^k \sum_{v \in S_2(B)} 1_{v \in T_i} = \sum_{i=1}^k s(T_i).$$

Since each  $T_i$  is a proper sub-block of  $B$ , we have that  $\ell(T_i) < (1 + \beta)s(B_i)$ ; hence,

$$\sum_{v \in S_2(B)} b_B(v) = \sum_{i=1}^k s(T_i) \geq \frac{1}{1 + \beta} \sum_{i=1}^k \ell(T_i).$$

We note that  $r(B) \leq s_2(B)$ . Then by Lemma 3.10, we have that

$$\begin{aligned} \text{rank}(P_{B, \tau_0}) &\geq s_2(B) - r(B) + \sum_{v \in S_2(B)} (b_B(v) - 1) \\ &\geq \sum_{v \in S_2(B)} (b_B(v) - 1) \\ &= \sum_{v \in S_2(B)} b_B(v) - s_2(B) \\ &\geq \frac{1}{1 + \beta} \sum_{i=1}^k \ell(T_i) - s_2(B) \\ &= \frac{\ell(B) - s_1(B)}{1 + \beta} - s_2(B) \\ &= \frac{\ell(B)}{1 + \beta} - \frac{1}{1 + \beta} s_1(B) - s_2(B) \\ &\geq \frac{\beta}{1 + \beta} s_1(B), \end{aligned}$$

since  $B$  is a critical block, we have  $\ell(B) \geq (1 + \beta)s(B)$ . Moreover, by Lemma 3.10,

$$\text{rank}(P_{B, \tau_0}) \geq s_2(B) - r(B) + \sum_{v \in S_2(B)} (b_B(v) - 1) = s_2(B) + \sum_{v \in R(B)} (b_B(v) - 2) \geq s_2(B).$$

The last inequality above is because  $b_B(v) \geq 2$  for every  $v \in R(B)$ . Thus,

$$\text{rank}(P_{B, \tau_0}) \geq \max \left\{ s_2(B), \frac{\beta}{1 + \beta} s_1(B) \right\}.$$

Let  $\lambda := s_1(B)/s(B)$ . Then,  $s_1(B) = \lambda s(B)$  and  $s_2(B) = (1 - \lambda)s(B)$  with  $\lambda \in (0, 1]$ . Therefore,

$$\text{rank}(P_{B, \tau_0}) \geq \max \left\{ 1 - \lambda, \frac{\beta}{1 + \beta} \lambda \right\} s(B).$$

Setting  $\lambda = \frac{1 + \beta}{1 + 2\beta}$  minimizes the maximum among the two quantities in the right-hand side (RHS). Hence,

$$\text{rank}(P_{B, \tau_0}) \geq \frac{\beta}{1 + 2\beta} s(B). \quad \square$$

### 3.2 Run-time of FLIP for MAX-CUT in the Complete Graph

In this subsection, we will show that a linear-sized improving sequence will improve the value of  $H(\tau)$  by some non-negligible amount with constant probability. Theorem 1.1 will follow from this result. First, we show a slight extension to a lemma of Reference [7].

**LEMMA 3.12.** *Let  $\phi > 0$  and  $X_1, \dots, X_m$  be independent random variables with density functions  $f_1, \dots, f_m : \mathbb{R} \rightarrow [0, \phi]$ . Let  $X := (x_1, \dots, x_m)^\top$  and  $\alpha_1, \dots, \alpha_k \in \mathbb{Z}^n$  be linearly independent vectors.*

Then for every  $\epsilon > 0$ ,

$$\Pr_X \left[ \langle \alpha_i, X \rangle > 0 \forall i \in [k] \text{ and } \sum_{i=1}^k \langle \alpha_i, X \rangle \leq \epsilon \right] \leq \frac{(\phi\epsilon)^k}{k!}.$$

The proof of Lemma 3.12 is presented in the appendix (see Appendix A). We will use Lemma 3.12 and the rank lower bound to bound the probability of existence of a starting configuration that is slow for critical blocks.

LEMMA 3.13. *Let  $B$  be a  $\beta$ -critical block where  $0 < \beta \leq 1$  with  $S_1(B) \neq \emptyset$ . Then the probability that there exists a starting configuration  $\tau_0 \in \{\pm 1\}^V$  such that  $B$  is  $\epsilon$ -slowly improving from  $\tau_0$  with respect to  $X$  is at most*

$$2^{s(B)} \frac{(2\phi\epsilon)^{\frac{\beta}{1+2\beta}s(B)}}{\left(\frac{\beta}{1+2\beta}s(B)\right)!}.$$

PROOF. Suppose that  $B$  is  $\epsilon$ -slowly improving from some  $\tau_0$  with respect to  $X$ . Since  $B$  is improving, we have that  $\langle M_{B,\tau_0}^t, X \rangle > 0$  for all  $t \in [\ell(B)]$ . Since  $B$  is  $\epsilon$ -slow, we have that  $\sum_{t=1}^{\ell(B)} \langle M_{B,\tau_0}^t, X \rangle \leq \epsilon$ . As every column of  $P_{B,\tau_0}$  is the sum of two columns of  $M_{B,\tau_0}$ , we have  $\langle P_{B,\tau_0}^C, X \rangle > 0$  for all pairs  $C \in \Gamma(B)$ . Moreover, every column of  $M_{B,\tau_0}$  contributes at most to two distinct columns of  $P_{B,\tau_0}$ . Hence  $\sum_{C \in \Gamma(B)} \langle P_{B,\tau_0}^C, X \rangle \leq 2\epsilon$ .

For  $\pi_f : S(B) \rightarrow \{\pm 1\}$  and  $\pi_c : V \setminus S(B) \rightarrow \{\pm 1\}$ , let us define  $\tau_{(\pi_f, \pi_c)} : V \rightarrow \{\pm 1\}$  as

$$\tau_{(\pi_f, \pi_c)}(u) := \begin{cases} \pi_f(u) & \text{if } u \in S(B) \text{ and} \\ \pi_c(u) & \text{if } u \in V \setminus S(B). \end{cases}$$

Let  $\mathcal{R}_{B,\tau,X}$  denote the event that  $B$  is  $\epsilon$ -slowly improving from the initial configuration  $\tau$  with respect to  $X$ . Then, by union bound, the required probability is at most

$$\sum_{\pi_f: S(B) \rightarrow \{\pm 1\}} \Pr_X \left[ \exists \pi_c : V \setminus S(B) \rightarrow \{\pm 1\} : \mathcal{R}_{B,\tau_{(\pi_f, \pi_c)},X} \right].$$

Now, consider a fixed choice of  $\pi_f : S(B) \rightarrow \{\pm 1\}$ . We would like to bound the following probability:

$$\Pr_X \left[ \exists \pi_c : V \setminus S(B) \rightarrow \{\pm 1\} : \mathcal{R}_{B,\tau_{(\pi_f, \pi_c)},X} \right].$$

Let us define an initial configuration  $\bar{\pi}_c : V \setminus S(B) \rightarrow \{\pm 1\}$  by  $\bar{\pi}_c(u) = 1$  for all  $u \in V \setminus S(B)$  and consider  $\sigma := \tau_{(\pi_f, \bar{\pi}_c)}$ . By Proposition 1, we have  $P_{L,\sigma} = P_{L,\tau_{(\pi_f, \pi_c)}}$  for every  $\pi_c : V \setminus S(B) \rightarrow \{\pm 1\}$ . Hence,

$$\Pr_X \left[ \exists \pi_c : V \setminus S(B) \rightarrow \{\pm 1\} : \mathcal{R}_{B,\tau_{(\pi_f, \pi_c)},X} \right] = \Pr_X \left[ \mathcal{R}_{B,\sigma,X} \right] \leq \frac{(2\phi\epsilon)^{\frac{\beta}{1+2\beta}s(B)}}{\left(\frac{\beta}{1+2\beta}s(B)\right)!}.$$

The last inequality above follows from Lemma 3.12 and Corollary 3.11. Hence, the required probability is at most

$$2^{s(B)} \frac{(2\phi\epsilon)^{\frac{\beta}{1+2\beta}s(B)}}{\left(\frac{\beta}{1+2\beta}s(B)\right)!}. \quad \square$$

For edge-weights  $X$ , let  $\mathcal{R}_X$  denote the event that there exist a sequence  $L$  of moves of length  $\lceil (1 + \beta)n \rceil$  and an initial configuration  $\tau_0 \in \{\pm 1\}^V$  such that  $L$  is  $\epsilon$ -slowly improving from  $\tau_0$  with

respect to  $X$ . The following two lemmas show that if  $\mathcal{R}_X$  happens, then there exists a starting configuration  $\tau_0 \in \{\pm 1\}^V$  and a  $\beta$ -critical block  $B$  such that  $B$  is  $\epsilon$ -slowly improving from  $\tau_0$  with respect to  $X$  and moreover,  $\ell(B) = \lceil (1 + \beta)s(B) \rceil$  and  $S_1(B) \neq \emptyset$ . Lemma 3.14 is similar to Lemma 4.1 in Angel, Bubeck, Peres, and Wei [1].

**LEMMA 3.14.** *Let  $L$  be a sequence of moves of length  $\lceil (1 + \beta)n \rceil$  for some  $\beta > 0$ . Then there exists a  $\beta$ -critical block  $B$  in  $L$  such that  $\ell(B) = \lceil (1 + \beta)s(B) \rceil$ .*

**PROOF.** Consider an inclusion-wise minimal block  $B$  in  $L$  such that  $\ell(B) \geq (1 + \beta)s(B)$ . We note that such a block exists, since  $\ell(L) = \lceil (1 + \beta)n \rceil \geq (1 + \beta)s(B)$ . By inclusion-wise minimality, we have that  $B$  is a  $\beta$ -critical block. Suppose  $\ell(B) \geq \lceil (1 + \beta)s(B) \rceil + 1$ . Then, consider the sub-block  $B' \subsetneq B$  obtained by removing the last vertex from  $B$ . For this sub-block, we have that  $s(B') \leq s(B)$  and  $\ell(B') = \ell(B) - 1 \geq \lceil (1 + \beta)s(B) \rceil \geq \lceil (1 + \beta)s(B') \rceil \geq (1 + \beta)s(B')$ , thus, contradicting the choice of  $B$ .  $\square$

The next lemma shows that  $\beta$ -critical blocks that are improving from an initial configuration with respect to some edge weights have at least one singleton vertex.

**LEMMA 3.15.** *Let  $B$  be a  $\beta$ -critical block where  $0 < \beta < 1$  with  $\ell(B) = \lceil (1 + \beta)s(B) \rceil$ . Suppose there exists an initial configuration  $\tau_0 \in \{\pm 1\}^V$  and edge weights  $X \in [-1, 1]^E$  such that  $B$  is improving from  $\tau_0$  with respect to  $X$ . Then,  $S_1(B) \neq \emptyset$ .*

**PROOF.** Suppose for the sake of contradiction that  $S_1(B) = \emptyset$ . Suppose  $2s(B) > \lceil (1 + \beta)s(B) \rceil$ . Since every vertex appears at least twice in the sequence  $B$ , we have

$$\ell(B) \geq 2s(B) > \lceil (1 + \beta)s(B) \rceil = \ell(B),$$

a contradiction. Therefore, we may assume that  $2s(B) = \lceil (1 + \beta)s(B) \rceil$  (i.e.,  $\ell(B) = 2s(B)$ ). Let  $S_{\text{even}}(B)$  and  $S_{\text{odd}}(B)$  denote the set of vertices that appear an even and odd number of times in the sequence  $B$ , respectively. Since  $B$  is improving from  $\tau_0$  with respect to  $X$ , it follows that  $S_{\text{odd}}(B) \neq \emptyset$  (otherwise, every vertex moves an even number of times in  $B$ , which means that the final configuration is the same as the initial configuration  $\tau_0$  and consequently, the sequence  $B$  would not have been improving). Now, we have

$$\ell(B) = \sum_{v \in S_{\text{even}}(B)} \#_B(v) + \sum_{v \in S_{\text{odd}}(B)} \#_B(v) \tag{2}$$

$$\geq 2|S_{\text{even}}(B)| + 3|S_{\text{odd}}(B)| \tag{3}$$

$$= 2s(B) + |S_{\text{odd}}(B)| \tag{4}$$

$$> \ell(B), \tag{5}$$

a contradiction, where Equation (3) follows from the assumption that  $S_1(B) = \emptyset$ , Equation (4) follows from  $|S(B)| = |S_{\text{even}}(B)| + |S_{\text{odd}}(B)|$ , and Equation (5) follows from  $\ell(B) = 2s(B)$  and  $|S_{\text{odd}}(B)| \geq 1$ .  $\square$

**LEMMA 3.16.** *Let  $G$  be the complete graph with  $|V| = n$  and for constants  $\eta > 0$ , and  $\beta \in (0, 1)$ , let*

$$\epsilon := e^{-\frac{2(1+\beta)}{\beta}} \left( \frac{\beta}{1+2\beta} \right) \phi^{-1} n^{-\left( \frac{1+2\beta+2\beta^2}{\beta} + \frac{\eta(1+2\beta)}{\beta} \right)}.$$

*Then, the probability (over the choices of  $X$ ) that there exists a sequence  $L$  of moves of length  $\lceil (1 + \beta)n \rceil$  and an initial configuration  $\tau_0 \in \{\pm 1\}^V$  such that  $L$  is  $\epsilon$ -slowly improving from  $\tau_0$  with respect to  $X$  is  $o(1)$ .*

PROOF. Let  $\mathcal{R}_X$  denote the event that there exist a sequence  $L$  of moves of length  $\lceil(1+\beta)n\rceil$  and an initial configuration  $\tau_0 \in \{\pm 1\}^V$  and such that  $L$  is  $\epsilon$ -slowly improving from  $\tau_0$  with respect to  $X$ . Lemmas 3.14 and 3.15 show that if  $\mathcal{R}_X$  happens, then there exists a starting configuration  $\tau_0 \in \{\pm 1\}^V$  and a  $\beta$ -critical block  $B$  such that  $B$  is  $\epsilon$ -slowly improving from  $\tau_0$  with respect to  $X$  and moreover,  $\ell(B) = \lceil(1+\beta)s(B)\rceil$  and  $S_1(B) \neq \emptyset$ .

Let  $\mathcal{B}$  be the set of  $\beta$ -critical blocks with  $\ell(B) \leq \lceil(1+\beta)s(B)\rceil$  and  $S_1(B) \neq \emptyset$ . Then, the probability of  $\mathcal{R}_X$  occurring is at most the sum over  $B \in \mathcal{B}$  of the probability that there exists  $\tau_0 \in \{\pm 1\}^V$  such that  $B$  is  $\epsilon$ -slowly improving from  $\tau_0$  with respect to  $X$ . By Lemma 3.13, this sum is at most

$$\sum_{B \in \mathcal{B}} 2^{s(B)} \frac{(2\phi\epsilon)^{\frac{\beta}{1+2\beta}s(B)}}{\left(\frac{\beta}{1+2\beta}s(B)\right)!} \leq \sum_{s=1}^n \binom{n}{s} s^{\lceil(1+\beta)s\rceil} 2^s \frac{(2\phi\epsilon)^{\frac{\beta}{1+2\beta}s}}{\left(\frac{\beta}{1+2\beta}s\right)!}.$$

The last inequality above is because  $\ell(B) = \lceil(1+\beta)s(B)\rceil$  and hence the number of possibilities for  $B \in \mathcal{B}$  with  $s(B) = s$  for a fixed  $s$  is at most  $\binom{n}{s} s^{\lceil(1+\beta)s\rceil}$ . Now, by using Stirling's approximation and the fact that  $\lceil(1+\beta)s\rceil \leq (1+\beta)s + 1$ , we have that the probability of  $\mathcal{R}_X$  occurring is at most

$$\sum_{s=1}^n \left(\frac{ne}{s}\right)^s s^{(1+\beta)s+1} 2^s \frac{(2\phi\epsilon)^{\frac{\beta}{1+2\beta}s}}{\left(\frac{\beta}{e(1+2\beta)}s\right)^{\frac{\beta}{1+2\beta}s}} \leq \sum_{s=1}^n s \left( e^{\frac{2(1+3\beta)}{1+2\beta}} \left(\frac{\beta}{1+2\beta}\right)^{-\frac{\beta}{1+2\beta}} n^{\frac{1+2\beta+2\beta^2}{1+2\beta}} (\phi\epsilon)^{\frac{\beta}{1+2\beta}} \right)^s.$$

The last inequality above is by using the fact that  $s \leq n$ . Now, for the choice of

$$\epsilon = e^{-\frac{2(1+3\beta)}{\beta}} \left(\frac{\beta}{1+2\beta}\right) \phi^{-1} n^{-\left(\frac{1+2\beta+2\beta^2}{\beta} + \frac{\eta(1+2\beta)}{\beta}\right)},$$

the above sum is an arithmetic-geometric sum. That is,

$$\Pr_X[\mathcal{R}_X] \leq \sum_{s=1}^n sn^{-\eta s} \leq \sum_{s=1}^{\infty} sn^{-\eta s} = \frac{n^{-\eta}}{(1-n^{-\eta})^2},$$

which tends to 0 as  $n \rightarrow \infty$ . □

We now restate and prove Theorem 1.1.

**THEOREM 1.1.** *Let  $G = (V, E)$  be the complete graph on  $n$  vertices, and suppose that the edge weights  $(X_e)_{e \in E}$  are independent random variables chosen according to a probability density function  $f_e : [-1, 1] \rightarrow [0, \phi]$  for some  $\phi > 0$ . Then, for every constant  $\eta > 0$ ,*

- (i) *with high probability every implementation of the FLIP method for MAX-CUT terminates in at most  $23,400\phi n^{3+(2+\sqrt{2})(\sqrt{2}+\eta)} = O(\phi n^{7.829+3.414\eta})$  steps.*
- (ii) *the expected number of steps in every implementation of the FLIP method for MAX-CUT is  $O(\phi n^{\frac{47}{6}+\eta})$ .*

PROOF.

- (i) We will use Lemma 3.16 with an optimal setting of  $\beta$ . We will derive this optimal setting at the end of the proof. For now, let us consider  $\beta \in (0, 1)$ . Let  $\mathcal{R}_X$  denote the event that an implementation of FLIP starting from some initial configuration  $\tau_0$  follows a sequence  $L$  of length at least

$$(1+\beta)e^{\frac{2(1+3\beta)}{\beta}} \left(\frac{\beta}{1+2\beta}\right)^{-1} \phi n^{\left(3+\frac{1+2\beta+2\beta^2}{\beta} + \frac{\eta(1+2\beta)}{\beta}\right)}. \quad (6)$$

Suppose  $\mathcal{R}_X$  happens. For  $1 \leq i \leq z := \ell(L)/(1+\beta)n$ , let  $L_i$  denote the block of  $L$  from time-step  $(i-1)(1+\beta)n+1$  to time-step  $i(1+\beta)n$  and let  $\tau_i$  denote the configuration before time-step  $(i-1)(1+\beta)n+1$ . We note that

$$z \geq e^{-\frac{2(1+3\beta)}{\beta}} \left( \frac{\beta}{1+2\beta} \right)^{-1} \phi n^{\left(2 + \frac{1+2\beta+2\beta^2}{\beta} + \frac{\eta(1+2\beta)}{\beta}\right)}. \quad (7)$$

For every  $i \in [z]$ , we have that  $\ell(L_i) = \lceil(1+\beta)n\rceil$  and  $L_i$  is an improving sequence from the initial configuration  $\tau_i$  with respect to  $X$ . We will now show that there exists  $i \in [z]$  such that  $L_i$  is an  $\epsilon$ -slowly improving sequence from the initial configuration  $\tau_i$  with respect to  $X$  for an appropriate choice of  $\epsilon$ .

For notational convenience, let  $h(L)$  denote the total improvement of  $H(\tau_0)$  from the initial configuration  $\tau_0$  by following the sequence of moves in  $L$ . Then,  $h(L) \leq n^2$ , since  $|X_e| \leq 1$  for every  $e \in E$ . Let  $h(L_i)$  denote the total improvement of  $H(\tau_i)$  from the initial configuration  $\tau_i$  by following the sequence of moves in  $L_i$ . Then,  $h(L) = \sum_{i=1}^z h(L_i)$ . Hence, there exists  $i \in [z]$  such that  $h(L_i)$  is at most

$$\frac{n^2}{z} \leq e^{-\frac{2(1+3\beta)}{\beta}} \left( \frac{\beta}{1+2\beta} \right) \phi^{-1} n^{-\left(\frac{1+2\beta+2\beta^2}{\beta} + \frac{\eta(1+2\beta)}{\beta}\right)}.$$

The inequality above is by the lower bound on  $z$  from Equation (7). Thus, there exists  $i \in [z]$  such that  $L_i$  is  $\epsilon$ -slowly improving from  $\tau_i$  with respect to  $X$ , where

$$\epsilon := e^{-\frac{2(1+3\beta)}{\beta}} \left( \frac{\beta}{1+2\beta} \right) \phi^{-1} n^{-\left(\frac{1+2\beta+2\beta^2}{\beta} + \frac{\eta(1+2\beta)}{\beta}\right)}.$$

The above argument implies that if  $\mathcal{R}_X$  happens, then there exists a sequence  $L'$  of moves of length  $\lceil(1+\beta)n\rceil$  and an initial configuration  $\sigma_0 \in \{\pm 1\}^V$  such that  $L'$  is  $\epsilon$ -slowly improving from  $\sigma_0$  with respect to  $X$ . By Lemma 3.16, the probability of the latter event is  $o(1)$  and hence the probability that  $\mathcal{R}_X$  happens is  $o(1)$ .

It remains to identify a setting of  $\beta$  that bounds the run-time. That is, we need a setting of  $\beta$  that minimizes the exponent of  $n$  in the RHS of Equation (6). The optimal choice is  $\beta = 1/\sqrt{2}$ . Thus, the probability that an implementation of FLIP starting from some initial configuration  $\tau_0$  follows a sequence  $L$  of length at least

$$23, 400 \phi n^{3+(2+\sqrt{2})(\sqrt{2}+\eta)}$$

is  $o(1)$ .

- (ii) We consider  $\beta = \max\{2/3+\eta/2, 1\} \in (2/3, 1)$ . Let  $L$  denote the sequence of moves executed by the algorithm and let  $T := \lfloor \ell(L)/(1+\beta)n \rfloor$ . If  $T \geq 1$ , then there exists a starting configuration  $\tau_0$  and a  $\beta$ -critical block  $B$  such that  $B$  is  $(n^2/T)$ -slowly improving from  $\tau_0$  with respect to  $X$  and moreover, by Lemmas 3.14 and 3.15,  $\ell(B) = \lceil(1+\beta)s(B)\rceil$  and  $S_1(B) \neq \emptyset$ .

For non-negative integers  $s$  and  $t$ , let  $R(s, t)$  denote the event that there exists a  $\beta$ -critical block  $B$  in  $L$  of length  $\lceil(1+\beta)s\rceil$  with  $s(B) = s$  and  $S_1(B) \neq \emptyset$  and a starting partition  $\tau_0$  such that  $B$  is  $(n^2/t)$ -slowly improving from  $\tau_0$  with respect to edge-weights  $X$ . Let

$$t_0 := \left\lceil e^{-\frac{2(1+3\beta)}{\beta}} \left( \frac{1+2\beta}{\beta} \right) \phi n^{4+2\beta+\frac{1}{\beta}} \right\rceil.$$

We have that

$$\mathbb{E}_X(T) \leq \sum_{t=1}^{\infty} \Pr_X(T \geq t) \leq t_0 + \sum_{t=t_0}^{\infty} \sum_{s=1}^n \Pr_X(R(s, t)).$$

We will show that the second term is  $O(t_0)$ .

We begin by considering the case of  $s \leq 3$ . For  $s = 1, 2$ , there are no  $\beta$ -critical blocks  $B$  with  $s(B) = s$  for which  $\ell(B) = \lceil(1 + \beta)s\rceil$ . For  $s = 3$ , an exhaustive search shows that for every  $\beta$ -critical block  $B$  with  $s(B) = 3$  from every starting configuration  $\tau_0$  with  $\ell(B) = \lceil 3(1 + \beta) \rceil = 6$ , the rank of the matrix  $P_{B, \tau_0}$  is at least 3. Thus,

$$\Pr_X(R(3, t)) \leq \binom{n}{3} 3^{\lceil(1+\beta)3\rceil} 2^3 \frac{\left(2\phi\left(\frac{n^2}{t}\right)\right)^3}{2} = O\left(n^3 \left(\frac{\phi n^2}{t}\right)^3\right).$$

Hence,

$$\sum_{t=t_0}^{\infty} \Pr_X(R(3, t)) = \int_{t_0}^{\infty} O\left(n^3 \left(\frac{\phi n^2}{t}\right)^3\right) dt = O\left(n^3 t_0 \left(\frac{\phi n^2}{t_0}\right)^3\right) = o(t_0).$$

The last equality above is by the choice of  $t_0$ .

We now consider the case of  $s \geq 4$ . By Lemma 3.13 and the fact that the number of critical blocks  $B$  with  $s(B) = s$  and  $\ell(B) = \lceil(1 + \beta)s(B)\rceil$  is at most  $\binom{n}{s} s^{\lceil(1+\beta)s\rceil}$ , we have that

$$\begin{aligned} \Pr_X(R(s, t)) &\leq \binom{n}{s} s^{\lceil(1+\beta)s\rceil} 2^s \frac{\left(2\phi\left(\frac{n^2}{t}\right)\right)^{\frac{\beta}{1+2\beta}s}}{\left(\frac{\beta}{1+2\beta}s\right)!} \\ &\leq \left(\frac{ne}{s}\right)^s s^{(1+\beta)s+1} 2^s \frac{\left(2\phi\left(\frac{n^2}{t}\right)\right)^{\frac{\beta}{1+2\beta}s}}{\left(\frac{\beta}{e(1+2\beta)}s\right)^{\frac{\beta}{1+2\beta}s}} \\ &\leq s \left( e^{\frac{2(1+3\beta)}{1+2\beta}} \left(\frac{\beta}{1+2\beta}\right)^{-\frac{\beta}{1+2\beta}} n s^{\frac{2\beta^2}{1+2\beta}} \left(\phi\left(\frac{n^2}{t}\right)\right)^{\frac{\beta}{1+2\beta}} \right)^s. \end{aligned}$$

Hence,  $\sum_{t=t_0}^{\infty} \Pr_X(R(s, t))$  is at most

$$\left( e^{\frac{2(1+3\beta)}{1+2\beta}} \left(\frac{\beta}{1+2\beta}\right)^{-\frac{\beta}{1+2\beta}} n s^{\frac{2\beta^2}{1+2\beta}} \left(\phi\left(\frac{n^2}{t_0}\right)\right)^{\frac{\beta}{1+2\beta}} \right)^s \left(\frac{st_0}{s - \frac{1+2\beta}{\beta}}\right) \left(\frac{1+2\beta}{\beta}\right).$$

With the above-mentioned choice of  $t_0$ , we obtain that

$$\sum_{t=t_0}^{\infty} \Pr_X(R(s, t)) \leq \left(\frac{s}{n}\right)^{\left(\frac{2\beta^2}{1+2\beta}\right)s} \left(\frac{st_0}{s - \frac{1+2\beta}{\beta}}\right) \left(\frac{1+2\beta}{\beta}\right).$$

Therefore,

$$\sum_{s=4}^n \sum_{t=t_0}^{\infty} \Pr_X(R(s, t)) \leq \left(\frac{1+2\beta}{\beta}\right) \sum_{s=4}^n \left(\frac{s}{n}\right)^{\left(\frac{2\beta^2}{1+2\beta}\right)s} \left(\frac{st_0}{s - \frac{1+2\beta}{\beta}}\right) = O(t_0) \sum_{s=4}^n \left(\frac{s}{n}\right)^{\left(\frac{2\beta^2}{1+2\beta}\right)s},$$

where the last equation is because  $\beta \in (0, 1)$  is a constant. Finally, we note that for every constant  $\beta$ , we have

$$\sum_{s=4}^n \left(\frac{s}{n}\right)^{\left(\frac{2\beta^2}{1+2\beta}\right)s} = O(1),$$

which is due to the fact that the function

$$f(n) := \sum_{s=4}^n \left(\frac{s}{n}\right)^{\left(\frac{2\beta^2}{1+2\beta}\right)s}$$

has a constant value for  $n = 4$ , and is decreasing for  $n \geq 4$  as we have

$$\frac{d}{dn} \left( \sum_{s=4}^n \left( \frac{s}{n} \right)^{\left( \frac{2\beta^2}{1+2\beta} \right)^s} \right) = \sum_{s=4}^n \frac{d}{dn} \left( \frac{s}{n} \right)^{\left( \frac{2\beta^2}{1+2\beta} \right)^s} = \sum_{s=4}^n -\frac{2\beta^2}{1+2\beta} \left( \frac{s}{n} \right)^{\frac{2\beta^2}{1+2\beta} s+1} < 0.$$

Hence,

$$\sum_{s=4}^n \sum_{\ell=t_0}^{\infty} \Pr_X(R(s, \ell)) = O(t_0).$$

We have thus shown that  $\mathbb{E}_X(T) = O(t_0)$  for every constant  $\beta \in (2/3, 1)$  and hence, the expected number of flips performed by the algorithm is

$$O(t_0 n) = O(\phi n^{5+2\beta+\frac{1}{\beta}}) = O(\phi n^{\frac{13}{2}+2\beta}).$$

By the choice of  $\beta = \max\{2/3 + \eta/2, 1\}$ , we see that the expected number of flips performed by the algorithm is

$$O(\phi n^{\frac{47}{6}+\eta}).$$

□

#### 4 SMOOTHED ANALYSIS OF FLIP FOR MAX-K-CUT

In this section, we prove Theorems 1.2 and 1.3. We begin with some notation. Let  $G = (V, E)$  be an arbitrary connected graph with  $n$  vertices and let  $X : E \rightarrow [-1, 1]$  be an edge-weight function. We will redefine some of the concepts from Section 3 as there are subtle differences between the same notions between the case of MAX-CUT and MAX- $k$ -CUT. For the sake of completeness, we state the complete definitions and prove all necessary details.

We recall a convenient formulation of the objective function for MAX- $k$ -CUT [9]. When considering MAX- $k$ -CUT, let  $\sigma(1), \dots, \sigma(k)$  be vectors defined as follows: take an equilateral simplex  $\Sigma_k$  in  $\mathbb{R}^{k-1}$  with vertices  $b_1, \dots, b_k$ . Let  $c_k := (b_1 + \dots + b_k)/k$  be the centroid of  $\Sigma_k$  and let  $\sigma(i) = b_i - c_k$ , for  $i \in [k]$ . Assume that  $\Sigma_k$  is scaled such that  $|\sigma(i)| = 1$  for  $i \in [k]$ . For example, MAX-3-CUT produces the vectors:

$$\sigma(1) := \frac{1}{\sqrt{6}}(-2, 1, 1), \sigma(2) := \frac{1}{\sqrt{6}}(1, -2, 1), \text{ and } \sigma(3) := \frac{1}{\sqrt{6}}(1, 1, -2).$$

*Remark 3.* If  $i, j \in [k]$ , then

$$\langle \sigma(i), \sigma(j) \rangle = \begin{cases} 1 & \text{if } i = j, \\ \frac{-1}{k-1} & \text{if } i \neq j. \end{cases}$$

We consider the space  $[k]^V$  of configurations that define a partition of the vertex set into  $k$  parts. For a configuration  $\tau \in [k]^V$ , we denote the part of  $v$  by  $\tau(v)$ . For a configuration  $\tau \in [k]^V$ , the weight of  $\tau$  is given by

$$\frac{k-1}{k} \sum_{uv \in E} X_{uv} (1 - \langle \sigma(\tau(u)), \sigma(\tau(v)) \rangle).$$

Let

$$H(\tau) := -\frac{k-1}{k} \sum_{uv \in E} X_{uv} \langle \sigma(\tau(u)), \sigma(\tau(v)) \rangle.$$

We observe that for constant  $k$ ,  $H(\tau)$  is a translation of the weight of  $\tau$  by some fraction of the total weight of all edges and hence, it suffices to work with  $H(\tau)$  henceforth.

We analyze the run-time of the FLIP method in the smoothed framework. We will denote the *move* of a vertex  $v \in V$  from part  $p \in [k]$  to part  $q \in [k] \setminus \{p\}$  as an ordered triple  $(v, p, q)$ . A move

$(v, p, q)$  is valid for a configuration  $\tau \in [k]^V$  if  $\tau(v) = p$  and  $q \neq p$ . We will need the notions of valid and improving sequences that we define now.

*Definition 4.1.* Let  $L$  be a sequence of moves,  $\tau_0 \in [k]^V$  be an initial configuration, and  $X \in [-1, 1]^E$  be the edge weights. We will denote the length of the sequence  $L$  by  $\ell(L)$ , the set of vertices appearing in the moves in  $L$  by  $S(L)$ , and  $s(L) := |S(L)|$ . For each  $v \in V$ , we will denote the number of times that the vertex  $v$  moves in  $L$  by  $\#_L(v)$ . We will denote the  $t$ th move of  $L$  by  $L(t) = (v_t, p_t, q_t)$ .

- (1) For each  $t \in [\ell(L)]$  such that  $L(t)$  is valid for  $\tau_{t-1}$ , we will denote  $\tau_t$  as the configuration obtained from  $\tau_{t-1}$  by setting  $\tau_t(u) := \tau_{t-1}(u)$  for every  $u \in V \setminus \{v_t\}$  and  $\tau_t(v_t) := q_t$ . If there exists  $t \in [\ell(L)]$  such that  $L(t)$  is invalid for  $\tau_{t-1}$ , then we say that  $L$  is invalid from  $\tau_0$ ; otherwise  $L$  is valid from  $\tau_0$ .
- (2) We say that  $L$  is improving from  $\tau_0$  with respect to  $X$  if  $L$  is valid from  $\tau_0$  and  $H(\tau_t) - H(\tau_{t-1}) > 0$  for all  $t \in [\ell(L)]$ . We say that  $L$  is  $\epsilon$ -slowly improving from  $\tau_0$  with respect to  $X$  if  $L$  is valid from  $\tau_0$  and  $H(\tau_t) - H(\tau_{t-1}) \in (0, \epsilon]$  for all  $t \in [\ell(L)]$ .

The notion of valid sequences is needed only for  $k \geq 3$  in the case of MAX- $k$ -CUT and was not necessary for MAX-CUT in the previous section. Moreover, we emphasize that the definition of  $\epsilon$ -slowly improving here is different from the one that we used in Section 3 for MAX-CUT. The improvement of  $H(\tau)$  can once again be written as the inner product between the edge-weight vector  $X$  and a suitable vector whose coordinates are in  $\{0, \pm 1\}$ . We formalize this with a convenient matrix below.

*Definition 4.2.* Let  $L$  be a valid sequence of moves from a configuration  $\tau_0 \in [k]^V$ . Let  $M_{L, \tau_0} \in \{0, \pm 1\}^{E \times \ell(L)}$  be a matrix with rows corresponding to the edges of  $G$ , columns corresponding to time-steps in the sequence  $L$ , and whose entries are given by

$$M_{L, \tau_0}[\{a, b\}, t] := \begin{cases} +1 & \text{if } a = v_t \text{ and } p_t = \tau_t(b), \\ & \text{or } b = v_t \text{ and } p_t = \tau_t(a), \\ -1 & \text{if } a = v_t \text{ and } q_t = \tau_t(b), \\ & \text{or } b = v_t \text{ and } q_t = \tau_t(a), \\ 0 & \text{otherwise,} \end{cases}$$

where  $\{a, b\} \in E$  and  $t \in [\ell(L)]$ . We will denote the  $t$ th column of  $M_{L, \tau_0}$  by  $M_{L, \tau_0}^t$ .

*Remark 4.* For a sequence  $L$  that is valid from an initial configuration  $\tau_0$ , we have  $H(\tau_t) - H(\tau_{t-1}) = \langle M_{L, \tau_0}^t, X \rangle$ .

Next, we generalize the notion of “pairs of immediate occurrences” used in Section 3 with the notion of cycles and cyclic vertices. We note that the following definitions do not depend on the initial configuration.

*Definition 4.3.* Let  $L$  be a sequence of moves.

- (1) A set of  $w$  moves  $\{(v_{t_1}, p_{t_1}, q_{t_1}), \dots, (v_{t_w}, p_{t_w}, q_{t_w})\}$  in  $L$  is a  $w$ -circuit over a vertex  $v \in S(L)$  if
  - (a)  $t_i < t_j$  for all  $i < j$ ,
  - (b)  $q_{t_i} = p_{t_{i+1}}$  for all  $i \in [w - 1]$ ,
  - (c)  $q_{t_w} = p_{t_1}$  and,
  - (d)  $v_{t_i} = v$  for all  $i \in [w]$ .

We will denote the time steps  $\{t_1, \dots, t_w\}$  of the  $w$ -circuit by  $T(C)$ .

- (2) A  $w$ -circuit is a  $w$ -cycle if it is inclusion-wise minimal.
- (3) A set  $C$  of moves in  $L$  is a *cycle* if it is a  $w$ -cycle for some  $w$ . Also, let  $t_{beg}(C) := \min(T(C))$  and  $t_{end}(C) := \max(T(C))$ . Let  $\Gamma(L)$  denote the set of all cycles in  $L$ .
- (4) A vertex  $v$  is called *cyclic* if there exists a cycle in  $\Gamma(L)$  that is over  $v$ . A vertex  $v$  is called *acyclic* if it is not cyclic. Let  $C(L)$  and  $A(L)$  denote the set of cyclic and acyclic vertices of  $L$ , respectively, and let  $c(L) := |C(L)|$  and  $a(L) := |A(L)|$ .

As mentioned above, the notion of cycles (circuits) can be viewed as a generalization of pairs of immediate occurrences used in Section 3. We note that every pair of immediate occurrences of a vertex is a 2-cycle (2-circuit). However, the converse does not hold, as the moves in a cycle (circuit) need not be consecutive moves of a vertex.

*Remark 5.* Let  $L$  be a sequence of moves and let  $C \in \Gamma(L)$  be a cycle over a vertex  $v$ . Then, every part is visited at most once by  $v$  in the cycle.

*Remark 6.* For a sequence  $L$  of moves, we have that  $\#_L(v) \leq k - 1$  for each vertex  $v \in A(L)$ .

We now define a suitable matrix that will nullify the influence of non-moving vertices.

*Definition 4.4.* Let  $L$  be a sequence of moves that is valid from an initial configuration  $\tau_0 \in [k]^V$ . Let  $P_{L, \tau_0} \in \{0, \pm 1\}^{E \times \Gamma(L)}$  be a matrix with rows corresponding to edges of  $G$ , columns corresponding to cycles in  $L$ , and whose entries are given by

$$P_{L, \tau_0}[\{a, b\}, C] := \sum_{(v_i, p_i, q_i) \in C} M_{L, \tau_0}[\{a, b\}, t_i],$$

where  $\{a, b\} \in E$  and  $C \in \Gamma(L)$ .

**PROPOSITION 2.** *For a sequence  $L$  of moves that is valid from an initial configuration  $\tau_0 \in [k]^V$ , if  $v \in V \setminus S(L)$ , then  $P_{L, \tau_0}[\{a, v\}, C] = 0$  for every  $C \in \Gamma(L)$  and  $\{a, v\} \in E$ .*

**PROOF.** Let  $C \in \Gamma(L)$  and  $\{a, v\} \in E$ . Since  $v$  is not in  $S(L)$ , it follows that  $C$  is not over  $v$ . If  $C$  is not over  $a$ , then  $M_{L, \tau_0}[\{a, v\}, t] = 0$  for every  $t \in T(C)$ , and hence  $P_{L, \tau_0}[\{a, v\}, C] = 0$ . Suppose  $C$  is a  $w$ -cycle over the vertex  $a$  and let  $C = \{(a, p_{t_1}, q_{t_1}), \dots, (a, p_{t_w}, q_{t_w})\}$ . If  $M_{L, \tau_0}[\{a, v\}, t_i] = 0$  for all  $i \in [w]$ , then the claim holds. Let  $t_{i_0}$  be the first instant in  $\{t_1, \dots, t_w\}$  such that  $M_{L, \tau_0}[\{a, v\}, t_{i_0}] \neq 0$ . Then, it follows that  $t_{i_0} \neq t_w$ . We have two cases:

- **Case 1.** Suppose  $t_{i_0} \neq t_1$ . Then  $M_{L, \tau_0}[\{a, v\}, t_{i_0}] = -1$  and  $M_{L, \tau_0}[\{a, v\}, t_{i_0+1}] = 1$ . Moreover,  $M_{L, \tau_0}[\{a, v\}, t_j] = 0$  for every  $j \in [w] \setminus \{i_0, i_0 + 1\}$ , since  $v$  does not move in  $L$ . Hence,  $P_{L, \tau_0}[\{a, v\}, C] = \sum_{i=1}^w M_{L, \tau_0}[\{a, v\}, t_i] = 0$ .
- **Case 2.** Suppose  $t_{i_0} = t_1$ . Then,  $M_{L, \tau_0}[\{a, v\}, t_1] = 1$ . Moreover,  $M_{L, \tau_0}[\{a, v\}, t_i] = 0$  for all  $1 < i < w$ , since  $q_{t_i} \neq p_{t_1}$  for all  $i \neq w$ . Finally, it follows that  $M_{L, \tau_0}[\{a, v\}, t_w] = -1$ , since  $v$  does not move between  $t_1$  and  $t_w$  in  $L$ . Hence,  $P_{L, \tau_0}[\{a, v\}, C] = \sum_{i=1}^w M_{L, \tau_0}[\{a, v\}, t_i] = 0$ .  $\square$

#### 4.1 Rank Lower Bounds for $P_{L, \tau_0}$

In this section, we show a lower bound on the rank of  $P_{L, \tau_0}$ . For this, we will again relate the rank of the matrix with the number of edges of a certain directed graph. This graph has certain properties that imply this relation. These properties, in fact, motivate the definition of the graph; hence, we define the properties in Definition 4.5 below before defining the graph.

The notion of  $L$ -good arcs below will help us identify non-zero entries in the matrix while the notion of  $L$ -neighbor-wise independence below will help us obtain large upper triangular submatrices.

*Definition 4.5.* Let  $L$  be a sequence of moves.

- (i) For  $u, v \in S(L)$ , we will call the ordered pair  $vu$  to be an  $L$ -good-arc if there exists a cycle  $C \in \Gamma(L)$  over  $v$  such that  $P_{L, \tau_0}[\{u, v\}, C] \neq 0$ . A directed graph  $H$  with node set  $S(L)$  is  $L$ -good if every arc in  $H$  is an  $L$ -good-arc.
- (ii) For a cyclic vertex  $v \in C(L)$  and a collection  $U \subseteq S(L)$  of vertices with  $m := |U|$ , the collection of ordered pairs  $\{vu : u \in U\}$  is an  $L$ -neighbor-wise independent set if there exists an ordering of  $U$ , say  $u_1, \dots, u_m$ , along with cycles  $C_1, \dots, C_m \in \Gamma(L)$  over  $v$  such that
  - (a)  $P_{L, \tau_0}[\{v, u_i\}, C_i] \neq 0$  and
  - (b)  $P_{L, \tau_0}[\{v, u_j\}, C_i] = 0$  for every  $j \in \{i + 1, \dots, m\}$ .

A directed graph  $H$  with node set  $S(L)$  is  $L$ -neighbor-wise independent if for every  $v \in S(L)$ , the collection  $\{vu : u \in \Delta_H^{out}(v)\}$  is an  $L$ -neighbor-wise independent set.

The next lemma is our key tool in obtaining a lower bound on the rank of  $P_{L, \tau_0}$ . We show that the rank is at least the number of edges in an  $L$ -good  $L$ -neighbor-wise independent directed acyclic graph. Although the definitions have changed mildly, the proof follows the same outline as that of Lemma 3.6 in Section 3.

LEMMA 4.6. *Let  $L$  be a valid sequence from an initial configuration  $\tau_0 \in [k]^V$ . Let  $H$  be an  $L$ -good  $L$ -neighbor-wise independent directed acyclic graph. Then,*

$$\text{rank}(P_{L, \tau_0}) \geq |E(H)|.$$

PROOF. Consider the submatrix  $B_H$  of  $P_{L, \tau_0}$  consisting of the rows corresponding to edges  $\{u, v\}$  for every arc  $vu \in E(H)$ . We will show that the matrix  $B_H$  has full row-rank by induction on  $|E(H)|$ . The base case of  $|E(H)| = 0$  is trivial.

For the induction step, we consider  $|E(H)| \geq 1$ . Suppose that there exist coefficients  $\mu_{\{u, v\}} \in \mathbb{R}$  for every  $uv \in E(H)$  such that

$$\sum_{uv \in E(H)} \mu_{\{u, v\}} P_{L, \tau_0}[\{u, v\}, C] = 0$$

for every cycle  $C \in \Gamma(L)$ . Since  $H$  is a directed acyclic graph with at least one arc, there exists a node  $v \in V(H)$  with  $|\delta_H^{out}(v)| \geq 1$  and  $|\delta_H^{in}(v)| = 0$ .

*Claim 3.* For every  $u \in \Delta_H^{out}(v)$ , the coefficient  $\mu_{\{v, u\}}$  is zero.

PROOF. Consider the ordering  $u_1, \dots, u_m$  of the vertices in  $\Delta_H^{out}(v)$  and cycles  $C_i \in \Gamma(L)$  satisfying Definition 4.5 (4.5). We want to show that  $\mu_{\{v, u_j\}} = 0$  for every  $j \in [m]$  by induction on  $j$ . We use induction on  $j \in [m + 1]$  to show that  $\mu_{\{v, u_i\}} = 0$  for every  $i < j$ .

The base case for  $j = 1$  is trivial. For the induction step, we consider  $j = j_0 + 1 \geq 2$ . We consider the columns of  $P$  corresponding to the cycle  $C_{j_0} \in \Gamma(L)$ . Since  $C_{j_0}$  is also over  $v$ , the only possible non-zero entries in this column among the chosen rows are in the rows corresponding to the edges  $\{v, u_1\}, \dots, \{v, u_m\}$ . Thus,

$$0 = \sum_{uv \in E(H)} \mu_{\{u, v\}} P_{L, \tau_0}[\{u, v\}, C_{j_0}] = \sum_{i=1}^m \mu_{\{v, u_i\}} P_{L, \tau_0}[\{v, u_i\}, C_{j_0}].$$

By induction hypothesis, we know that  $\mu_{\{v, u_i\}} = 0$  for every  $i < j_0$ . Thus,

$$0 = \sum_{i=j_0}^m \mu_{\{v, u_i\}} P_{L, \tau_0}[\{v, u_i\}, C_{j_0}].$$

Moreover, by the choice of  $C_{j_0}$ , we also have that

$$P_{L, \tau_0}[\{v, u_{j_0}\}, C_{j_0}] \neq 0, \text{ and } P_{L, \tau_0}[\{v, u_i\}, C_{j_0}] = 0 \forall i \in \{j_0 + 1, \dots, m\}.$$

Consequently, we obtain that  $\mu_{\{v, u_{j_0}\}} = 0$ . We have thus concluded the inductive statement that  $\mu_{\{v, u_i\}} = 0$  for every  $i < j_0 + 1$ .  $\square$

As a consequence of the claim, we have that the matrix  $B_H$  has full row-rank if and only if the matrix  $B_{H'}$  obtained from the graph  $H' := H - \delta_H^{out}(v)$  has full row-rank. We note that the graph  $H'$  is also an  $L$ -good  $L$ -neighbor-wise independent directed acyclic graph with  $|E(H')| < |E(H)|$ . Thus, by induction hypothesis, the matrix  $B_{H'}$  has full row-rank. Hence, the matrix  $B_H$  also has full row-rank.  $\square$

We now use Lemma 4.6 to show that the rank of the matrix  $P_{L, \tau_0}$  is at least half the number of cyclic vertices in  $L$  provided that  $L$  is an improving sequence from some initial configuration.

LEMMA 4.7. *Let  $L$  be an improving sequence from an initial configuration  $\tau_0 \in [k]^V$  with respect to some edge weights  $X \in [-1, 1]^E$ . Then,*

$$\text{rank}(P_{L, \tau_0}) \geq \frac{1}{2}c(L).$$

PROOF. We will use the following claim to construct an  $L$ -good  $L$ -neighbor-wise independent directed acyclic graph.

Claim 4. For every vertex  $v \in C(L)$  and for every cycle  $C \in \Gamma(L)$  that is over  $v$ , there exists an edge  $\{u, v\} \in E$  such that  $P_{L, \tau_0}[\{u, v\}, C] \neq 0$ .

PROOF. For contradiction, suppose that for all edges  $\{u, v\} \in E$ , we have  $P_{L, \tau_0}[\{u, v\}, C] = 0$ . We note that for all  $e \in E$ , if  $v$  is not an end-vertex of  $e$ , then  $M_{L, \tau_0}[e, t] = 0$  for every  $t \in T(C)$ . Moreover, for all  $e \in E$  with  $v$  being an end-vertex of  $e$ , we have that

$$\sum_{t \in T(C)} M_{L, \tau_0}[e, t] = P_{L, \tau_0}[e, C] = 0.$$

Hence, for every  $e \in E$ , we have that

$$\sum_{t \in T(C)} M_{L, \tau_0}[e, t] = 0.$$

This implies that  $\sum_{t \in T(C)} \langle M_{L, \tau_0}^t, Y \rangle = 0$  for all  $Y \in [-1, 1]^E$ . However, since  $L$  is an improving sequence from  $\tau_0$  with respect to  $X \in [-1, 1]^E$ , it follows that  $\langle M_{L, \tau_0}^t, X \rangle > 0$  for all  $t \in [\ell(L)]$ . In particular,  $\sum_{t \in T(C)} \langle M_{L, \tau_0}^t, X \rangle > 0$ , a contradiction.  $\square$

Now, we construct an  $L$ -good  $L$ -neighbor-wise independent graph  $H$  over the node set  $S(L)$  as follows: For every  $v \in C(L)$ , pick an arbitrary  $u \in V \setminus \{v\}$  such that  $P_{L, \tau_0}[\{u, v\}, C] \neq 0$  (which is guaranteed to exist by the above claim) and add the arc  $vu$  to  $H$ . The resulting graph  $H$  is  $L$ -good by construction. It is trivially  $L$ -neighbor-wise independent, since each node has out-degree at most one. Moreover,  $|E(H)| = c(L)$  and the directed cycles in  $H$  are node-disjoint.

Finally, we obtain an  $L$ -good  $L$ -neighbor-wise independent directed acyclic graph  $H'$  by removing one arc from each directed cycle in  $H$ . Since  $|E(H)| = c(L)$  and the directed cycles in  $H$  are node-disjoint, it follows that  $|E(H')| \geq \frac{1}{2}c(L)$ . The lemma now follows by applying Lemma 4.6 to  $H'$ .  $\square$

Next, we will improve the rank lower bound from Lemma 4.7 in complete graphs for MAX-3-CUT. We need a few additional definitions. The following definition will also be useful for the quasi-polynomial time analysis for MAX- $k$ -CUT.

*Definition 4.8.* Let  $L$  be a sequence of moves. A *block* is a continuous subsequence of  $L$ . For a block  $L'$  of  $L$ , we will denote the set of time-steps of the moves of  $L'$  in  $L$  by  $T(L')$ . A maximal block of  $L$  consisting only of cyclic vertices of  $L$  is called a cyclic block. Likewise, a maximal block of  $L$  consisting only of acyclic vertices of  $L$  is called an acyclic block.

We note that a sequence  $L$  can be partitioned into alternating cyclic and acyclic blocks.

*4.1.1 Improving Rank Lower Bounds for MAX-3-CUT in the Complete Graph.* We now focus on the case when  $k = 3$  and  $G$  is the complete graph. For  $k = 3$ , the cycles of interest are 2-cycles and 3-cycles. We are able to sharpen the rank lower bound with a better understanding of how the cycles interact with the cyclic blocks. We need the following definitions to characterize these interactions.

*Definition 4.9.* A cycle  $C \in \Gamma(L)$  over a vertex  $v \in V$  is *leaping* if the time-steps in  $T(C)$  belong to at least two distinct cyclic blocks of  $L$ . A leaping 3-cycle  $\{(v_t, p_t, q_t), (v_{t'}, p_{t'}, q_{t'}), (v_{t''}, p_{t''}, q_{t''})\} \in \Gamma(L)$  is called *tricky* if

- (1) the time-steps  $t, t', t''$  belong to distinct cyclic blocks of  $L$  and
- (2) the set of acyclic vertices of  $L$  that appear between  $t$  and  $t'$  is the same as those that appear between  $t'$  and  $t''$  (i.e.,  $A(L) \cap S(L[t, t']) = A(L) \cap S(L[t', t''])$ ).

The following two lemmas summarize the structure of cyclic blocks and leaping cycles.

**LEMMA 4.10.** *Let  $L$  be a valid sequence from some initial configuration  $\tau_0 \in [3]^V$  and let  $t_1 < t_2 < t_3$  be the time-steps of three occurrences of a vertex  $v$  in  $L$  such that  $t_1, t_2, t_3$  belong to different cyclic blocks. Then, there exists a cycle  $C \in \Gamma(L)$  over  $v$  such that*

- (i)  $C$  is a leaping cycle and
- (ii)  $t \in \{t_1, t_1 + 1, \dots, t'_3\}$  for every  $t \in T(C)$ , where  $t'_3$  is the last occurrence of  $v$  in the same cyclic block as that of  $t_3$ .

**PROOF.** Let the time-steps  $t_1, t_2$  and  $t_3$  be in cyclic blocks  $B_1, B_2$ , and  $B_3$ , respectively. Let  $t'_1, t'_2$  and  $t'_3$  denote the last occurrence of  $v$  in  $B_1, B_2$ , and  $B_3$ , respectively. Without loss of generality, suppose that the vertex  $v$  moves from part 1 to part 2 at time-step  $t'_1$  (i.e.,  $p_{t'_1} = 1$  and  $q_{t'_1} = 2$ ). We consider the following three cases:

- **Case 1.** There exists a time-step  $t$  such that  $t'_1 \leq t \leq t'_3$  and  $L(t) = (v, 2, 1)$ . Then, the cycle  $C \in \Gamma(L)$  given by  $\{(v, 1, 2), (v, 2, 1)\}$  with  $T(C) = \{t'_1, t\}$  is the desired cycle.
- **Case 2.** There does not exist a time-step  $t$  such that  $t'_1 \leq t \leq t'_3$  and  $L(t) = (v, 2, 1)$ , but there exists a time-step  $t$  such that  $t'_1 \leq t \leq t'_3$  and  $L(t) = (v, 3, 1)$ . Since  $L$  is a valid sequence from an initial configuration  $\tau_0$ , there exists a time-step  $t'$  such that  $t'_1 \leq t' \leq t$  and  $L(t') = (v, 2, 3)$ . Hence, the cycle  $C \in \Gamma(L)$  given by  $\{(v, 1, 2), (v, 2, 3), (v, 3, 1)\}$  with  $T(C) = \{t'_1, t', t\}$  is the desired cycle.
- **Case 3.** There does not exist a time-step  $t$  such that  $t'_1 \leq t \leq t'_3$  and  $L(t) = (v, 2, 1)$ , and there does not exist a time-step  $t$  such that  $t'_1 \leq t \leq t'_3$  and  $L(t) = (v, 3, 1)$ . Let  $t$  be the first occurrence of  $v$  in  $B_3$ . Since  $L$  is a valid sequence for an initial configuration  $\tau_0$ , we have that  $\{L(t'_2), L(t)\} = \{(v, 2, 3), (v, 3, 2)\}$ . Hence, the cycle  $C \in \Gamma(L)$  given by  $\{(v, 2, 3), (v, 3, 2)\}$  with  $T(C) = \{t'_2, t\}$  is the desired cycle.  $\square$

LEMMA 4.11. *Let  $G = (V, E)$  be the complete graph and let  $L$  be a valid sequence from some initial configuration  $\tau_0 \in [3]^V$ . Suppose  $C \in \Gamma(L)$  is a non-tricky leaping cycle over a vertex  $v$ . Then there exists a vertex  $u \in A(L) \cap S(L[t_{beg}(C), t_{end}(C)])$  such that the arc  $vu$  is an  $L$ -good-arc.*

PROOF. We consider the following three cases:

- **Case 1.** Suppose that  $C$  is a 2-cycle and  $T(C) = \{t_1, t_2\}$ . Without loss of generality, suppose that  $C = \{(v, 1, 2), (v, 2, 1)\}$ . Since  $C$  is a leaping cycle, there exists a vertex  $u \in A(L) \cap S(L[t_{beg}(C), t_{end}(C)])$ . Let  $p, q \in [3]$  denote the parts of  $u$  at time-steps  $t_1$  and  $t_2$ , respectively. Since  $u \in A(L)$ , we have that  $p \neq q$ . Therefore,  $\{p, q\} \in \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$ . Hence,  $P_{L, \tau_0}[\{u, v\}, C] \in \{\pm 1, \pm 2\}$  and it is non-zero.
- **Case 2.** Suppose that  $C$  is a 3-cycle and the time-steps in  $T(C) = \{t_1, t_2, t_3\}$  belong to exactly two distinct cyclic blocks of  $L$ . Without loss of generality, suppose that  $C = \{(v, 1, 2), (v, 2, 3), (v, 3, 1)\}$  and that  $t_1, t_2$  belong to the same cyclic block, which is different from that of  $t_3$ . We note that there exists  $u \in A(L) \cap S(L[t_2, t_3])$ . Let  $p, q \in [3]$  denote the parts of  $u$  at time-steps  $t_2$  and  $t_3$ , respectively. Since  $u \in A(L)$ , we have that  $p \neq q$ . Therefore,  $\{p, q\} \in \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$ . Hence,  $P_{L, \tau_0}[\{u, v\}, C] \in \{\pm 1, \pm 2\}$  and it is non-zero.
- **Case 3.** Suppose that  $C$  is a 3-cycle and the time-steps in  $T(C) = \{t_1, t_2, t_3\}$  belong to exactly three distinct cyclic blocks of  $L$ . Without loss of generality, suppose that  $C = \{(v, 1, 2), (v, 2, 3), (v, 3, 1)\}$ . Since  $C$  is non-tricky, there exists  $u \in A(L)$  such that  $u$  is in exactly one of  $S(L[t_1, t_2])$  and  $S(L[t_2, t_3])$ . Without loss of generality, suppose that  $u \in S(L[t_2, t_3]) \setminus S(L[t_1, t_2])$ . Let  $p, q \in [3]$  denote the position of  $u$  at time-steps  $t_2$  and  $t_3$ , respectively. Since  $u \in A(L)$ , we have that  $p \neq q$ . Therefore,  $\{p, q\} \in \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$ . Hence,  $P_{L, \tau_0}[\{u, v\}, C] \in \{\pm 1, \pm 2\}$  and it is non-zero.  $\square$

We now show that for every cyclic vertex  $v$  in a sequence  $L$ , we have a large number of  $L$ -good arcs whose tail is  $v$  that also form an  $L$ -neighbor-wise independent set. This fact will be useful in constructing a large  $L$ -good  $L$ -neighbor-wise independent directed acyclic graph, which will in turn improve the rank using Lemma 4.6. We emphasize that our proof of this fact will crucially use the fact that the graph  $G$  is complete. For a sequence  $L$  of moves, we will denote the number of cyclic blocks in which  $v$  occurs as  $b_L(v)$ .

LEMMA 4.12. *Let  $G$  be the complete graph, let  $L$  be a valid sequence from some initial configuration  $\tau_0 \in [3]^V$ , and let  $v \in C(L)$ . Then, there exists a collection of  $k \geq \lfloor \frac{1}{2} \lfloor \frac{1}{3} (b_L(v) - 1) \rfloor \rfloor$  vertices  $u_1, \dots, u_k$  such that*

- (i)  $u_1, \dots, u_k \in A(L)$ ,
- (ii)  $vu_i$  is an  $L$ -good-arc for every  $i \in [k]$ , and
- (iii) the set  $\{vu_1, \dots, vu_k\}$  is an  $L$ -neighbor-wise independent set.

PROOF. Let the cyclic blocks that contain  $v$  be  $B_1, \dots, B_{b_L(v)}$ . Let  $R := \lfloor \frac{1}{3} (b_L(v) - 1) \rfloor$ . For  $0 \leq r \leq R$ , let  $\mathbf{B}_r := \{B_{3r+1}, B_{3r+2}, B_{3r+3}, B_{3r+4}\}$ . We note that the last block in  $\mathbf{B}_r$  and the first block in  $\mathbf{B}_{r+1}$  coincide for every  $0 \leq r < R$ . Let  $\mathbf{A}_r$  be the set of three acyclic blocks in  $L$  that appear between  $B_{3r+1}$  and  $B_{3r+2}$ , between  $B_{3r+2}$  and  $B_{3r+3}$  and between  $B_{3r+3}$  and  $B_{3r+4}$ . For simplicity, we let  $S(\mathbf{A}_r) := \cup_{A \in \mathbf{A}_r} S(A)$  denote the set of vertices occurring in the blocks in  $\mathbf{A}_r$ , and let  $T(\mathbf{A}_r) := \cup_{A \in \mathbf{A}_r} T(A)$  denote the set of time-steps of the moves of blocks in  $\mathbf{A}_r$  as they appear in  $L$ . By definition, we have that  $T(\mathbf{A}_r) \cap T(\mathbf{A}_{r'}) = \emptyset$  for every distinct  $r, r' \in [R]$ .

Let us consider an  $r \in [R]$ . By Lemma 4.10, there exist leaping cycles  $C_r$  and  $C'_r$  in  $\Gamma(L)$  over  $v$  such that  $T(C_r) \subseteq T(B_{3r+1}) \cup T(B_{3r+2}) \cup T(B_{3r+3})$  and  $T(C'_r) \subseteq T(B_{3r+2}) \cup T(B_{3r+3}) \cup T(B_{3r+4})$ .

*Claim 5.* The cycles  $C_r$  and  $C'_r$  cannot both be tricky cycles.

PROOF. For the sake of contradiction suppose that both  $C_r$  and  $C'_r$  are tricky cycles. Let  $T(C_r) = \{t_1, t_2, t_3\}$  and  $T(C'_r) = \{t'_1, t'_2, t'_3\}$  for some  $t_1 < t_2 < t_3$  and  $t'_1 < t'_2 < t'_3$ . By the choice of  $C_r$  and  $C'_r$ , we must have  $t_i \in T(B_{3r+i})$  and  $t'_i \in T(B_{3r+i+1})$  for  $i \in \{1, 2, 3\}$ . Therefore,  $A(L) \cap S(L[t_2, t_3]) = A(L) \cap S(L[t'_1, t'_2])$ , and since  $C_r$  and  $C'_r$  are tricky, it follows that  $A(L) \cap S(L[t_1, t_2]) = A(L) \cap S(L[t_2, t_3])$  and  $A(L) \cap S(L[t'_1, t'_2]) = A(L) \cap S(L[t'_2, t'_3])$ . Hence,  $A(L) \cap S(L[t_1, t_2]) = A(L) \cap S(L[t_2, t_3]) = A(L) \cap S(L[t'_2, t'_3])$ . Since neither of these sets are empty, there must exist a  $u \in A(L)$  that appears at least 3 times in the sequence. However, this contradicts the fact that for all  $u \in A(L)$ , we have  $\#_L(u) \leq 2$ .  $\square$

Let  $\hat{C}_r := C_r$  if  $C_r$  is non-tricky, and  $\hat{C}_r := C'_r$  otherwise. It follows from Claim 5 that  $\hat{C}_r$  is a non-tricky cycle. Therefore, by Lemma 4.11, there exists a vertex  $u_r \in A(L)$  that appears between two cyclic blocks in  $\mathbf{B}_r$  such that  $\nu u_r$  is an  $L$ -good-arc.

We have shown that for each  $r \in [R]$ , there exists a vertex  $u_r \in S(\mathbf{A}_r)$  that appears between two cyclic blocks in  $\mathbf{B}_r$  such that  $\nu u_r$  is an  $L$ -good-arc. Let  $U := \{u_1, \dots, u_R\}$ . We note that the vertices  $u_1, \dots, u_R$  may not be distinct and consequently, we may not be able to obtain a large number of  $L$ -good-arcs while constructing the needed  $L$ -good  $L$ -neighbor-wise independent directed acyclic graph. Even if they are distinct, we need an ordering of them that satisfies the  $L$ -neighbor-wise independent property. The following claim handles these two issues.

*Claim 6.* There exists a subset of  $k \geq R/2$  distinct elements in  $U$  along with an ordering  $u_{r_1}, \dots, u_{r_k}$  of these elements such that  $u_{r_j} \notin S(\mathbf{A}_{r_i})$  for every  $i, j \in [k]$  with  $i < j$ .

PROOF. We will construct a sequence  $w_1, \dots, w_k$  such that

- (i)  $k \geq R/2$ ,
- (ii)  $w_i \in U$  for every  $i \in [k]$ ,
- (iii)  $w_i \neq w_j$  for every  $1 \leq i < j \leq k$ ,
- (iv) for every  $1 \leq i < j \leq k$ , we have that  $w_i \notin S(\mathbf{A}_t)$ , where  $t$  is any index in  $[R]$  such that  $w_j = u_t$ .

We show that such a sequence translates into the sequence  $u_{r_1}, \dots, u_{r_k}$  required in the claim. Conditions (i), (ii), and (iii) imply that the elements  $w_1, \dots, w_k$  are indeed a subset of  $k \geq R/2$  distinct elements of  $U$ . Let us denote the sequence  $w_k, \dots, w_1$  by  $u_{r_1}, \dots, u_{r_k}$ , where  $r_i$  is the least index of  $w_{k-i+1}$  in  $u_1, \dots, u_R$ . Substituting  $w_i = u_{r_{k-i+1}}$  and  $w_j = u_{r_{k-j+1}}$  for all  $1 \leq i < j \leq k$  in condition (iv) results in  $u_{r_{k-i+1}} \notin S(\mathbf{A}_{r_{k-j+1}})$  for every  $1 \leq i < j \leq k$ . Thus, re-indexing produces  $u_{r_j} \notin S(\mathbf{A}_{r_i})$  for every  $i, j \in [k]$  with  $i < j$  as desired.

To construct a sequence  $w_1, \dots, w_k$  satisfying the conditions (i)–(iv), we consider the procedure in Figure 2. The input satisfies the required conditions as  $u_r \in S(\mathbf{A}_r)$  for every  $r \in [R]$  by construction of  $u_1, \dots, u_R$ , and  $u_r$  appears in at most two of the  $\mathbf{A}_i$ 's, since  $\#_L(u_r) \leq 2$  and  $T(\mathbf{A}_r) \cap T(\mathbf{A}_{r'}) = \emptyset$  for every distinct  $r, r' \in [R]$ .

We first show that the procedure always terminates. We note that before the execution of step 2(d), we have  $r \in I$  because of step 2(b). Moreover, from  $w_k = u_r \in S(\mathbf{A}_r)$  it follows that  $r$  will be removed from  $I$  after the execution of step 2(d). Hence, the size of  $I$  decreases by at least one after each iteration of the while loop.

Next, we prove conditions (i)–(iv). Suppose the procedure terminates with the sequence  $w_1, \dots, w_k$ .

- (i) Since each  $u_r$  appears in at most two  $S(\mathbf{A}_i)$ 's, each execution of step 2(d) decreases the size of  $I$  by at most two. Therefore, the while loop iterates at least  $R/2$  times. Hence,  $k \geq R/2$ .
- (ii) By step 2(c), we have that  $w_i \in U$  for every  $i \in [k]$ .
- (iii) From step 2(b) to 2(d) it follows that  $w_i \neq w_j$  for  $1 \leq i < j \leq k$ .

---

**Procedure**


---

**Input:** Sets  $A_1, \dots, A_R$  and elements  $u_1, \dots, u_R$  with  $u_r \in S(A_r)$  for every  $r \in [R]$ , such that  $u_r$  appears in at most two of the  $S(A_i)$ 's (i.e. at most one other than  $A_r$ ).

- (1) Initialize  $k \leftarrow 0, I \leftarrow [R]$ .
  - (2) While ( $I \neq \emptyset$ ):
    - (a)  $k \leftarrow k + 1$ .
    - (b)  $r \leftarrow$  an arbitrary element in  $I$ .
    - (c)  $w_k \leftarrow u_r$ .
    - (d)  $I \leftarrow I \setminus \{i \in [R] : w_k \in S(A_i)\}$ .
  - (3) Return  $(w_1, \dots, w_k)$ .
- 

Fig. 2. Procedure for Claim 6.

- (iv) Let  $i \in [k]$ . Consider the  $i$ th iteration of the while loop. After step 2(d) refines the set  $I$ , we have that  $w_i \notin S(A_r)$  for all indices  $r \in I$  that are chosen in iterations after the  $i$ th iteration. Thus, if  $t$  is the index of  $w_j$  in  $u_1, \dots, u_R$  for some  $j \in [k]$  where  $j > i$ , then  $w_i \notin S(A_t)$ .  $\square$

Consider the subset of  $k \geq \lceil \frac{1}{2} \lfloor \frac{1}{3}(b_L(v) - 1) \rfloor \rceil$  distinct elements  $U' = \{u_{r_1}, \dots, u_{r_k}\} \subseteq U$  from Claim 6. Conditions (4.12) and (4.12) desired in the lemma trivially hold, since  $U' \subseteq U \subseteq A(L)$  and  $vu$  is an  $L$ -good arc for all  $u \in U$ . We recall that the cycles  $C_{r_1}, \dots, C_{r_k}$  are disjoint and  $P_{L, \tau_0}[\{v, u_{r_i}\}, C_{r_i}] \neq 0$  for all  $i \in [k]$ . Then, since  $u_{r_i} \notin S(A_{r_j})$  for every  $j \in \{i + 1, \dots, k\}$ , it follows that  $P_{L, \tau_0}[\{v, u_{r_i}\}, C_{r_j}] = 0$  for all  $j \in \{i + 1, \dots, k\}$ . Hence, Condition (4.12) holds.  $\square$

We now have all the tools necessary to show our improved rank lower bound.

**LEMMA 4.13.** *Let  $G = (V, E)$  be the complete graph and let  $L$  be a valid sequence from some initial configuration  $\tau_0 \in [3]^V$ . Then  $\text{rank}(P_{L, \tau_0}) \geq \frac{1}{6} \sum_{v \in C(L)} (b_L(v) - 3)$ .*

**PROOF.** For each  $v \in C(L)$ , let  $R(v) := \lceil \frac{1}{2} \lfloor \frac{1}{3}(b_L(v) - 1) \rfloor \rceil$ . We construct an  $L$ -good  $L$ -neighbor-wise independent directed acyclic graph  $H$  over the vertex set  $S(L)$  as follows: for each  $v \in C(L)$ , add the arcs  $vu_{r_1}, \dots, vu_{R(v)}$ , where  $u_{r_1}, \dots, u_{R(v)}$  is the collection of vertices guaranteed to exist by Lemma 4.12. Then it follows that  $H$  is  $L$ -good and  $L$ -neighbor-wise independent. Moreover, we note that all arcs in  $E(H)$  have heads in  $A(L)$  and tails in  $C(L)$ . Hence,  $H$  does not contain any directed cycles. Therefore,  $H$  is an  $L$ -good  $L$ -neighbor-wise independent directed acyclic graph. For a lower bound on the number of arcs in  $H$ , we have

$$\begin{aligned}
 |E(H)| &= \sum_{v \in C(L)} R(v) \\
 &= \sum_{v \in C(L)} \left\lceil \frac{1}{2} \left\lfloor \frac{1}{3}(b_L(v) - 1) \right\rfloor \right\rceil \\
 &= \sum_{v \in C(L)} \left\lceil \frac{1}{2} \left( \frac{1}{3}(b_L(v) - 1) - \frac{2}{3} \right) \right\rceil \\
 &\geq \sum_{v \in C(L)} \frac{1}{2} \left( \frac{b_L(v) - 3}{3} \right) \\
 &= \frac{1}{6} \sum_{v \in C(L)} (b_L(v) - 3). \quad \square
 \end{aligned}$$

Finally, we combine the results of Lemmas 4.7 and 4.13 to get one final rank bound. We define a block  $B$  in  $L$  to be *2-critical* if  $\ell(B) \geq 3s(B)$  and  $\ell(B') < 3s(B')$  for every block  $B'$  that is strictly contained in  $B$ .

**COROLLARY 4.14.** *Let  $G = (V, E)$  be the complete graph and let  $B$  be an improving sequence from some initial configuration  $\tau_0 \in [3]^V$  with respect to some edge weights  $X \in [-1, 1]^E$ . If  $B$  is a 2-critical block, then  $\text{rank}(P_{B, \tau_0}) \geq \frac{1}{32}s(B)$ .*

**PROOF.** Suppose that  $B$  has no acyclic vertices, then

$$\text{rank}(P_{B, \tau_0}) \geq \frac{1}{2}c(B) = \frac{1}{2}s(B),$$

and we are done. So, we may assume that  $B$  has at least one acyclic block. Let  $B_1, \dots, B_k$  denote the cyclic blocks of  $B$  and for a vertex  $v$ , let  $1_{v \in B_i}$  denote the indicator function for whether the vertex  $v$  appears in  $B_i$ . Then,

$$\sum_{v \in C(B)} b_B(v) = \sum_{v \in C(B)} \sum_{i=1}^k 1_{v \in B_i} = \sum_{i=1}^k \sum_{v \in C(B)} 1_{v \in B_i} = \sum_{i=1}^k s(B_i).$$

Since each  $B_i$  is a proper sub-block of  $B$ , we have that  $\ell(B_i) < 3s(B_i)$ ; hence,

$$\sum_{v \in C(B)} b_B(v) = \sum_{i=1}^k s(B_i) \geq \frac{1}{3} \sum_{i=1}^k \ell(B_i).$$

By Lemma 4.13, we have that

$$\text{rank}(P_{B, \tau_0}) \geq \frac{1}{6} \sum_{v \in C(B)} (b_B(v) - 3) \quad (8)$$

$$= \frac{1}{6} \left( \sum_{v \in C(B)} b_B(v) - 3c(B) \right) \quad (9)$$

$$\geq \frac{1}{6} \left( \frac{1}{3} \sum_{i=1}^k \ell(B_i) - 3c(B) \right) \quad (10)$$

$$\geq \frac{1}{6} \left( \frac{\ell(B) - 2a(B)}{3} - 3c(B) \right) \quad (11)$$

$$\geq \frac{1}{6} \left( s(B) - \frac{2}{3}a(B) - 3c(B) \right) \quad (12)$$

$$= \frac{1}{18}a(B) - \frac{1}{3}c(B), \quad (13)$$

where Equation (11) follows from the fact that acyclic vertices of  $B$  appear at most twice in  $B$ , inequality Equation (12) follows from the fact that  $B$  is a 2-critical block, and Equation (13) follows from  $s(B) = a(B) + c(B)$ . Thus, by the above inequality and Lemma 4.7, we have

$$\text{rank}(P_{B, \tau_0}) \geq \max \left\{ \frac{1}{2}c(B), \frac{1}{18}a(B) - \frac{1}{3}c(B) \right\}.$$

Let  $\lambda := a(B)/s(B)$ . Then,  $a(B) = \lambda s(B)$  and  $c(B) = (1 - \lambda)s(B)$ . Thus,

$$\text{rank}(P_{B, \tau_0}) \geq \max \left\{ \frac{1}{2}(1 - \lambda), \frac{1}{18}\lambda - \frac{1}{3}(1 - \lambda) \right\} s(B) \geq \frac{1}{32}s(B).$$

□

## 4.2 Run-time of FLIP for MAX-3-CUT in the Complete Graph

In this subsection, we will show that an improving sequence will improve the value of  $H(\tau)$  by some non-negligible amount with constant probability. Theorem 1.2 will follow from this result. First, we recall a lemma of Reference [7] below that closely resembles Lemma 3.12.

LEMMA 4.15. [7] *Let  $\phi > 0$  and  $X_1, \dots, X_m$  be independent random variables with density functions  $f_1, \dots, f_m : \mathbb{R} \rightarrow [0, \phi]$ , and let  $X := (x_1, \dots, x_m)^\top$ . Let  $\alpha_1, \dots, \alpha_k \in \mathbb{Z}^n$  be linearly independent vectors. Then for every  $\epsilon > 0$ ,*

$$\Pr_X \left[ \langle \alpha_i, X \rangle \in (0, \epsilon] \forall i \in [k] \right] \leq (\phi\epsilon)^k.$$

We note that the event of interest in Lemma 3.12, namely, all inner products being non-negative and having sum at most  $\epsilon$  implies the event of interest in Lemma 4.15, namely, all inner products being in the interval  $(0, \epsilon]$ . Consequently, the upper bound that we get for the latter event is weaker than the upper bound of Lemma 3.12. We will use Lemma 4.15 and the rank lower bound to bound the probability of the existence of a bad starting configuration for critical blocks.

LEMMA 4.16. *Let  $G = (V, E)$  be the complete graph and let  $B$  be a 2-critical block. Then the probability that there exists a starting configuration  $\tau_0 \in [3]^V$  such that  $B$  is  $\epsilon$ -slowly improving from  $\tau_0$  with respect to  $X$  is at most*

$$3^{s(B)} (3\phi\epsilon)^{\frac{1}{32}s(B)}.$$

PROOF. Let  $\bar{I}_{B, \tau_0}$  denote the event that there exist edge weights  $Y$  such that  $B$  is an improving sequence from  $\tau_0$  with respect to  $Y$ . Suppose that  $B$  is  $\epsilon$ -slowly improving from some  $\tau_0$  with respect to  $X$ . Since  $B$  is improving from  $\tau_0$  with respect to  $X$ , it follows that  $\langle M_{B, \tau_0}^t, X \rangle \in (0, \epsilon]$  for all  $t \in [\ell(B)]$ . Since every column of  $P_{B, \tau_0}$  is the sum of at most three columns of  $M_{B, \tau_0}$ , we have that  $\langle P_{B, \tau_0}^C, X \rangle \in (0, 3\epsilon]$  for all  $C \in \Gamma(B)$ . Hence, the required probability is at most the probability that there exists a starting configuration  $\tau_0 \in [3]^V$  such that  $\bar{I}_{B, \tau_0}$  holds and  $\langle P_{B, \tau_0}^C, X \rangle \in (0, 3\epsilon]$  for all  $C \in \Gamma(B)$ . Then, by union bound, the required probability is at most the sum over all  $v \in S(B)$  and  $\tau_0(v) \in [3]$  of the probability that for all  $u \in V \setminus S(B)$  there exists  $\tau_0(u) \in [3]$  such that  $\bar{I}_{B, \tau_0}$  holds and  $\langle P_{B, \tau_0}^C, X \rangle \in (0, 3\epsilon]$  for all  $C \in \Gamma(B)$ .

For  $\pi_f : S(B) \rightarrow [3]$  and  $\pi_c : V \setminus S(B) \rightarrow [3]$ , let us define  $\tau_{(\pi_f, \pi_c)} : V \rightarrow [3]$  as

$$\tau_{(\pi_f, \pi_c)}(u) := \begin{cases} \pi_f(u) & \text{if } u \in S(B) \text{ and} \\ \pi_c(u) & \text{if } u \in V \setminus S(B). \end{cases}$$

We will now bound the following probability for a fixed choice of  $\pi_f : S(B) \rightarrow [3]$  and then take a union bound over the choices of the initial configuration for the vertices in  $S(B)$ :

$$\Pr_X \left[ \exists \pi_c : V \setminus S(B) \rightarrow [3] : \bar{I}_{B, \tau_{(\pi_f, \pi_c)}} \text{ and } \langle P_{B, \tau_{(\pi_f, \pi_c)}}^C, X \rangle \in (0, 3\epsilon] \forall C \in \Gamma(B) \right].$$

Let us define an initial configuration

$$\sigma_0(u) := \begin{cases} \pi_f(u) & \text{if } u \in S(B) \text{ and} \\ 1 & \text{if } u \in V \setminus S(B). \end{cases}$$

By Proposition 2, we have that  $P_{B, \sigma_0} = P_{B, \tau_{(\pi_f, \pi_c)}}$  for all  $\pi_c : V \setminus S(B) \rightarrow [3]$ . Hence,

$$\begin{aligned} & \Pr_X \left[ \exists \pi_c : V \setminus S(B) \rightarrow [3] : \bar{I}_{B, \tau_{(\pi_f, \pi_c)}} \text{ and } \langle P_{B, \tau_{(\pi_f, \pi_c)}}^C, X \rangle \in (0, 3\epsilon] \forall C \in \Gamma(B) \right] \\ &= \Pr_X \left[ \exists \pi_c : V \setminus S(B) \rightarrow [3] : \bar{I}_{B, \tau_{(\pi_f, \pi_c)}} \text{ and } \langle P_{B, \sigma_0}^C, X \rangle \in (0, 3\epsilon] \forall C \in \Gamma(B) \right]. \end{aligned}$$

Now, we bound the RHS probability. If there exists  $\pi_c : V \setminus S(B) \rightarrow [3]$  such that the sequence  $B$  is improving from  $\tau_{(\pi_f, \pi_c)}$  with respect to some edge weights  $Y$ , then by Corollary 4.14, the rank of  $P_{B, \tau_{(\pi_f, \pi_c)}}$  is at least  $s(B)/32$ . Moreover, we know that  $P_{B, \sigma_0} = P_{B, \tau_{(\pi_f, \pi_c)}}$  and hence the rank of  $P_{B, \sigma_0}$  is at least  $s(B)/32$ . Therefore, using Lemma 4.15, the RHS probability is at most

$$(3\phi\epsilon)^{\text{rank}(P_{B, \sigma_0})} \leq (3\phi\epsilon)^{\frac{1}{32}s(B)}.$$

Hence, the probability required in the lemma is at most

$$3^{s(B)}(3\phi\epsilon)^{\frac{1}{32}s(B)},$$

as the number of possible initial configurations for the vertices that move in  $B$  is at most  $3^{s(B)}$ .  $\square$

**LEMMA 4.17.** *Let  $G$  be the complete graph and let  $\epsilon = \phi^{-1}n^{-(96+\eta)}$  for a constant  $\eta > 0$ . Then, the probability (over the choices of  $X$ ) that there exists a sequence  $L$  of moves of length  $3n$  and an initial configuration  $\tau_0 \in [3]^V$  such that  $L$  is  $\epsilon$ -slowly improving from  $\tau_0$  with respect to  $X$  is  $o(1)$ .*

**PROOF.** Let  $\mathcal{R}_X$  denote the event that there exists a sequence  $L$  of moves of length  $3n$  and an initial configuration  $\tau_0 \in [3]^V$  such that  $L$  is  $\epsilon$ -slowly improving from  $\tau_0$  with respect to  $X$ . We note that every sequence  $L$  of length  $3n$  contains a 2-critical block  $B$  whose length is  $3s(B)$ —the proof of this fact is similar to the proof of Lemma 3.14. Therefore, if the event  $\mathcal{R}_X$  happens, then there exists a 2-critical block  $B$  with  $\ell(B) = 3s(B)$  and an initial configuration  $\tau_0 \in [3]^V$  such that  $B$  is  $\epsilon$ -slowly improving from  $\tau_0$  with respect to  $X$ . Hence, the probability that  $\mathcal{R}_X$  occurs (over the choices of  $X$ ) is at most

$$\begin{aligned} & \sum_{\substack{B: B \text{ is critical,} \\ \ell(B) \leq 3n}} \Pr_X \left[ \exists \sigma_0 : B \text{ is } \epsilon\text{-slowly improving from } \tau_0 \text{ with respect to } X \right] \\ & \leq \sum_{\substack{B: B \text{ is critical,} \\ \ell(B) \leq 3n}} 3^{s(B)}(3\phi\epsilon)^{\frac{1}{32}s(B)} \quad (\text{by Lemma 4.16}) \\ & \leq \sum_{s=1}^n n^{3s} 3^s (3\phi\epsilon)^{\frac{1}{32}s} \\ & \leq \sum_{s=1}^n \left( Cn^3 \phi^{\frac{1}{32}} \epsilon^{\frac{1}{32}} \right)^s \end{aligned}$$

for some universal constant  $C > 0$ . The second inequality above is because a 2-critical block could have length at most  $3s$ . Therefore, for  $\epsilon = \phi^{-1}n^{-(96+\eta)}$  the sum tends to 0 as  $n \rightarrow \infty$ .  $\square$

The proof of Theorem 1.2(i) follows from Lemma 4.17 similar to the proof of Theorem 1.1(i) that follows from Lemma 3.16. In this case, we consider the event that an implementation of FLIP produces a sequence of moves from some initial configuration that has length greater than  $\phi n^{99+\eta}$  for any constant  $\eta > 0$ . This event implies that there exists a sequence of length  $3n$  that is  $\epsilon$ -slowly improving from some initial configuration, where  $\epsilon = \phi^{-1}n^{-(96+\eta)}$ . Finally, we note that the probability of such an event is  $o(1)$  by Lemma 4.17.

*Remark 7.* The run-time analysis using the above techniques can be improved by replacing the notion of 2-critical blocks with  $\beta$ -critical blocks and optimizing the value of  $\beta$  similar to the ideas in Section 3. This gives a run-time bound of  $O(\phi n^{90.81+\eta})$  with high probability. We avoid writing a proof of this improved bound in the interest of simplicity. We now prove Theorem 1.2(ii). We restate it below for the reader's convenience.

LEMMA 4.18. *Let  $G = (V, E)$  be the complete graph on  $n$  vertices, and suppose that the edge weights  $(X_e)_{e \in E}$  are independent random variables chosen according to a probability density function  $f_e : [-1, 1] \rightarrow [0, \phi]$  for some  $\phi > 0$ . Then, the expected number of steps in every implementation of the FLIP method for MAX-3-CUT is  $O(\phi n^{99})$ .*

PROOF. Let  $L$  denote the sequence of moves executed by the algorithm and let  $T := \lfloor \ell(L)/3n \rfloor$ . If  $T \geq 1$ , then there exists a starting configuration  $\tau_0$  and a 2-critical block  $B$  in  $L$  with  $\ell(B) = 3s(B)$  such that  $B$  is  $(n/T)$ -slowly improving from  $\tau_0$  with respect to  $X$ .

For non-negative integers  $s$  and  $t$ , let  $R(s, t)$  denote the event that there exists a 2-critical block  $B$  in  $L$  of length  $3s(B)$  and a starting partition  $\tau_0$  such that  $B$  is  $(n/t)$ -slowly improving from  $\tau_0$  with respect to edge weights  $X$ . Let  $\zeta := \lceil 4^{32} \phi n^{97} \rceil$  and  $t_i := i\zeta$  for  $i \geq 2$ . We recall that  $\ell(L) \leq 3^n$ , since the algorithm never cycles back to the same configuration twice. Therefore,

$$\begin{aligned} \mathbb{E}_X(T) &\leq \sum_{t=1}^{3^n} \Pr_X(T \geq t) \\ &\leq 2\zeta + \sum_{i=2}^{3^n} \sum_{t=t_i}^{t_{i+1}-1} \Pr_X(T \geq t) \\ &\leq 2\zeta + \sum_{i=2}^{3^n} \sum_{t=t_i}^{t_{i+1}-1} \Pr_X(T \geq t_i) \\ &\leq 2\zeta + \sum_{i=2}^{3^n} \sum_{t=t_i}^{t_{i+1}-1} \sum_{s=1}^n \Pr_X(R(s, t_i)). \end{aligned}$$

We will show that the second term in the RHS above is  $O(\zeta n)$ .

By Lemma 4.16 and the fact that the number of 2-critical blocks  $B$  with  $\ell(B) = 3s$  is at most  $n^{3s}$ , we have that

$$\Pr_X(R(s, t_i)) \leq 3^s n^{3s} \left( \frac{3\phi n}{t_i} \right)^{\frac{s}{32}} \leq \left( 4n^3 \left( \frac{\phi n}{i\zeta} \right)^{\frac{1}{32}} \right)^s \leq \left( \frac{1}{i} \right)^{\frac{s}{32}}.$$

The last inequality above is by the choice of  $\zeta$ . Hence,

$$\begin{aligned} \sum_{s=1}^n \Pr_X(R(s, t_i)) &\leq \sum_{s=1}^n \left( \frac{1}{i} \right)^{\frac{s}{32}} \\ &\leq \sum_{s'=0}^{\lceil n/32 \rceil} \sum_{s''=0}^{31} \left( \frac{1}{i} \right)^{s'+s''/32} \\ &\leq \sum_{s'=0}^{\infty} \sum_{s''=0}^{31} \left( \frac{1}{i} \right)^{s'+s''/32} \\ &\leq \sum_{s'=1}^{\infty} 32 \left( \frac{1}{i} \right)^{s'} \\ &\leq 32 \left( \sum_{s'=0}^{\infty} \frac{1}{i^{s'}} - 1 \right) \\ &= \frac{32}{i-1}. \end{aligned}$$

Therefore,

$$\sum_{i=2}^{3^n} \sum_{t=t_i}^{t_{i+1}-1} \sum_{s=1}^n \Pr_X(R(s, t_i)) \leq \sum_{i=2}^{3^n} \sum_{t=t_i}^{t_{i+1}-1} \frac{32}{i-1} = \sum_{i=2}^{3^n} \frac{32\zeta}{i-1} = 32\zeta O(\log(3^n)) = O(\zeta n).$$

We have thus shown that  $\mathbb{E}_X(T) = O(\zeta n)$  and hence, the expected number of flips performed by the algorithm is  $O(\zeta n^2) = O(\phi n^{99})$ .  $\square$

### 4.3 Run-time of FLIP for MAX- $k$ -CUT in Arbitrary Graphs

In this subsection, we turn our attention to FLIP for MAX- $k$ -CUT in arbitrary graphs and prove Theorem 1.3. We will again use Lemma 4.15. We note that it is not immediately clear that a result similar to Lemma 4.13 holds for arbitrary graphs. The proof technique for Lemma 4.13 fails for arbitrary graphs, since the acyclic vertices between two cyclic blocks for a vertex  $v$  may not be in the neighborhood of  $v$ . Instead, we show that each sequence of sufficiently large length must have a block that has a large fraction of cyclic vertices.

*Definition 4.19.* The surplus of a sequence  $L$  is

$$z(L) := \ell(L) - \sum_{v \in A(L)} \#_L(v) - c(L),$$

and the maximum surplus over all blocks of length  $t$  in  $L$  is

$$m_L(t) := \max\{z(B) : B \text{ is a block of } L \text{ with } \ell(B) = t\}.$$

The following lemma can be seen as a generalization of Lemma 5.1 from Angel, Bubeck, Peres, and Wei [1].

LEMMA 4.20. *Suppose that  $\alpha > 1$  and  $L$  is a sequence of length  $\alpha n$ . Then there exists a block  $B$  in  $L$  such that*

$$\frac{c(B)}{\ell(B)} \geq \frac{\alpha - k + 1}{(2k - 1)\alpha \log(\alpha n)}.$$

PROOF. We will use the following claim.

*Claim 7.* If a block  $B$  is the concatenation of blocks  $B_1$  and  $B_2$ , then

$$z(B) \leq z(B_1) + z(B_2) + (2k - 1)c(B).$$

PROOF. We recall Remark 6, which states that an acyclic vertex appears at most  $k - 1$  times in a sequence. Hence, in the worst case, a cyclic vertex  $v \in C(B)$  in  $B$  is an acyclic vertex in  $B_1$  and  $B_2$  and appears  $k - 1$  times in each block.  $\square$

Let  $\delta := \max\{c(B)/\ell(B) : B \text{ is a block in } L\}$ . Therefore, for every block  $B$  in  $L$ , we have  $c(B) \leq \delta \ell(B)$ . Then,

$$m_L(t) \leq 2m_L(t/2) + (2k - 1)\delta t$$

for all  $t > 0$ . Since  $m(1) = 0$ , we can bound the sequence above by

$$m_L(\alpha n) \leq (2k - 1)\delta \alpha n \log(\alpha n). \quad (14)$$

Since

$$\sum_{v \in A(B)} \#_B(v) + c(B) \leq (k - 1)s(B)$$

for all blocks  $B$  in  $L$  and  $s(L) \leq n$ , it follows that

$$m_L(\alpha n) \geq \alpha n - (k - 1)s(L) \geq (\alpha - k + 1)n. \quad (15)$$

Therefore, combining Equations (14) and (15), we get that

$$\delta \geq \frac{\alpha - k + 1}{(2k - 1)\alpha \log(\alpha n)},$$

which concludes the proof.  $\square$

*Definition 4.21.* A block  $B$  is  $\alpha$ -cyclic if

$$c(B) \geq \frac{\alpha - k + 1}{(2k - 1)\alpha \log(\alpha n)} \ell(B).$$

LEMMA 4.22. *Let  $G = (V, E)$  be a graph and let  $B$  be an  $\alpha$ -cyclic block. Then the probability that there exists a starting configuration  $\tau_0 \in [k]^V$  such that  $B$  is  $\epsilon$ -slowly improving from  $\tau_0$  with respect to  $X$  is at most*

$$k^{\ell(B)} (k\phi\epsilon)^{\frac{\alpha-k+1}{2(2k-1)\alpha \log(\alpha n)} \ell(B)}.$$

PROOF. For a starting configuration  $\tau_0 \in [k]^V$ , let  $\mathcal{I}_{B, \tau_0}$  denote the event that there exist edge weights  $Y$  such that  $B$  is an improving sequence from  $\tau_0$  with respect to  $Y$ . Suppose that  $B$  is  $\epsilon$ -slowly improving from some  $\tau_0 \in [k]^V$  with respect to  $X$ . Since  $B$  is improving from  $\tau_0$  with respect to  $X$ , it follows that  $\langle M_{B, \tau_0}^t, X \rangle \in (0, \epsilon]$  for all  $t \in [\ell(B)]$ . Since every column of  $P_{B, \tau_0}$  is the sum of at most  $k$  columns of  $M_{B, \tau_0}$ , we have that  $\langle P_{B, \tau_0}^C, X \rangle \in (0, k\epsilon]$  for all  $C \in \Gamma(B)$ . Hence, the required probability is at most the probability that there exists a starting configuration  $\tau_0 \in [k]^V$  such that  $\mathcal{I}_{B, \tau_0}$  holds and  $\langle P_{B, \tau_0}^C, X \rangle \in (0, k\epsilon] \forall C \in \Gamma(B)$ . Then, by union bound, the required probability is at most the sum over all  $v \in S(B)$  and  $\tau_0(v) \in [k]$  of the probability that for all  $u \in V \setminus S(B)$  there exists  $\tau_0(u) \in [k]$  such that  $\mathcal{I}_{B, \tau_0}$  holds and  $\langle P_{B, \tau_0}^C, X \rangle \in (0, k\epsilon] \forall C \in \Gamma(B)$ .

For  $\pi_f : S(B) \rightarrow [k]$  and  $\pi_c : V \setminus S(B) \rightarrow [k]$ , let us define  $\tau_{(\pi_f, \pi_c)} : V \rightarrow [k]$  as

$$\tau_{(\pi_f, \pi_c)}(u) := \begin{cases} \pi_f(u) & \text{if } u \in S(B) \text{ and} \\ \pi_c(u) & \text{if } u \in V \setminus S(B). \end{cases}$$

Now, consider a fixed choice of  $\pi_f$  and  $\pi_c$ . We would like to bound the following probability:

$$\Pr_X \left[ \exists \pi_c : V \setminus S(B) \rightarrow [k] : \mathcal{I}_{B, \tau_{(\pi_f, \pi_c)}} \text{ and } \langle P_{B, \tau_{(\pi_f, \pi_c)}}^C, X \rangle \in (0, k\epsilon] \forall C \in \Gamma(B) \right].$$

Let us define an initial configuration

$$\sigma_0(u) := \begin{cases} \pi_f(u) & \text{if } u \in S(B) \text{ and} \\ 1 & \text{if } u \in V \setminus S(B). \end{cases}$$

By Proposition 2, we have that  $P_{B, \sigma_0} = P_{B, \tau_{(\pi_f, \pi_c)}}$  for every  $\pi_c : V \setminus S(B) \rightarrow [k]$ . Hence,

$$\begin{aligned} & \Pr_X \left[ \exists \pi_c : V \setminus S(B) \rightarrow [k] : \mathcal{I}_{B, \tau_{(\pi_f, \pi_c)}} \text{ and } \langle P_{B, \tau_{(\pi_f, \pi_c)}}^C, X \rangle \in (0, k\epsilon] \forall C \in \Gamma(B) \right] \\ &= \Pr_X \left[ \exists \pi_c : V \setminus S(B) \rightarrow [k] : \mathcal{I}_{B, \tau_{(\pi_f, \pi_c)}} \text{ and } \langle P_{B, \sigma_0}^C, X \rangle \in (0, k\epsilon] \forall C \in \Gamma(B) \right]. \end{aligned}$$

Now, we bound the RHS probability. If there exists  $\pi_c : V \setminus S(B) \rightarrow [k]$  such that the sequence  $B$  is improving from  $\tau_{(\pi_f, \pi_c)}$  with respect to some edge weights  $Y$ , then by Lemma 4.7, the rank of  $P_{B, \tau_{(\pi_f, \pi_c)}}$  is at least  $c(B)/2$ . Moreover, we know that  $P_{B, \sigma_0} = P_{B, \tau_{(\pi_f, \pi_c)}}$ , hence the rank of  $P_{B, \sigma_0}$  is at least  $c(B)/2$ . Since  $B$  is an  $\alpha$ -cyclic block, we have that  $c(B) \geq (\alpha - k + 1)\ell(B)/(2k - 1)\alpha \log(\alpha n)$  and hence the rank of  $P_{B, \sigma_0}$  is at least  $(\alpha - k + 1)\ell(B)/(2(2k - 1)\alpha \log(\alpha n))$ . Therefore, using Lemma 4.15, the RHS probability is at most

$$(k\phi\epsilon)^{\text{rank}(P_{B, \sigma_0})} \leq (k\phi\epsilon)^{\frac{1}{2} \frac{\alpha-k+1}{(2k-1)\alpha \log(\alpha n)} \ell(B)}.$$

Hence, the probability required in the lemma is at most

$$k^{s(B)} (k\phi\epsilon)^{\frac{1}{2}} \cdot \frac{\alpha-k+1}{(2k-1)\alpha \log(\alpha n)} \ell(B).$$

as the number of possible initial configurations for the vertices that move in  $B$  is at most  $k^{s(B)}$ . The lemma now follows, since  $s(B) \leq \ell(B)$  and  $k \geq 1$ .  $\square$

The next lemma follows from Lemmas 4.22 and 4.20 similar to the proof of Lemma 4.17.

**LEMMA 4.23.** *Let  $G$  be a graph and let  $\epsilon = \phi^{-1}n^{-(2k-1)k \log(kn)+\eta}$  for a constant  $\eta > 0$ . Then, the probability (over the choices of  $X$ ) that there exists a sequence  $L$  of moves of length  $kn$  and an initial configuration  $\tau_0 \in [k]^V$  such that  $L$  is  $\epsilon$ -slowly improving from  $\tau_0$  with respect to  $X$  is  $o(1)$ .*

**PROOF.** Let  $\mathcal{R}_X$  denote the event that there exists a sequence  $L$  of moves of length  $kn$  and an initial configuration  $\tau_0 \in [k]^V$  such that  $L$  is  $\epsilon$ -slowly improving from  $\tau_0$  with respect to  $X$ . It follows from Lemma 4.20 that every sequence  $L$  of length  $kn$  contains a  $k$ -cyclic block. Therefore, if the event  $\mathcal{R}_X$  happens, then there exists a  $k$ -cyclic block  $B$  of length at most  $kn$  and an initial configuration  $\tau_0 \in [k]^V$  such that  $B$  is  $\epsilon$ -slowly improving from  $\tau_0$  with respect to  $X$ . Hence, the probability that  $\mathcal{R}_X$  occurs (over the choices of  $X$ ) is at most

$$\begin{aligned} & \sum_{\substack{B: B \text{ is } k\text{-cyclic,} \\ \ell(B) \leq kn}} \Pr_X \left[ \exists \sigma_0 \in [k]^V : B \text{ is } \epsilon\text{-slowly improving from } \tau_0 \text{ with respect to } X \right] \\ & \leq \sum_{\substack{B: B \text{ is } k\text{-cyclic,} \\ \ell(B) \leq kn}} k^{\ell(B)} (k\phi\epsilon)^{\frac{1}{2(2k-1)k \log(kn)}} \ell(B) \quad (\text{by Lemma 4.22}) \\ & \leq \sum_{\ell=1}^{kn} n^\ell k^\ell (k\phi\epsilon)^{\frac{\ell}{2(2k-1)k \log(kn)}} \\ & \leq \sum_{\ell=1}^{kn} \left( k^2 n \phi^{\frac{1}{2(2k-1)k \log(kn)}} \epsilon^{\frac{1}{2(2k-1)k \log(kn)}} \right)^\ell, \end{aligned}$$

where the second inequality follows from the fact that there are at most  $n^\ell$  blocks of length  $\ell$ . Therefore, for constant  $k$  and  $\epsilon = \phi^{-1}n^{-(2(2k-1)k \log(kn)+\eta)}$ , the sum tends to 0 as  $n \rightarrow \infty$ .  $\square$

The proof of Theorem 1.3(i) follows from Lemma 4.23 similar to the proof of Theorem 1.1(i) that follows from Lemma 3.16. In this case, we consider the event that an implementation of FLIP produces a sequence of moves from some initial configuration that has length greater than  $\phi n^{2(2k-1)k \log(kn)+3+\eta}$  for any constant  $\eta > 0$ . This event implies that there exists a sequence of length  $kn$  that is  $\epsilon$ -slowly improving from some initial configuration, where  $\epsilon = \phi^{-1}n^{-(2(2k-1)k \log(kn)+\eta)}$ . Finally, we note that the probability of such an event is  $o(1)$  by Lemma 4.23. We now prove Theorem 1.3(ii). We restate it below for the reader's convenience.

**LEMMA 4.24.** *Let  $G = (V, E)$  be an arbitrary graph on  $n$  vertices, and suppose that the edge weights  $(X_e)_{e \in E}$  are independent random variables chosen according to a probability density function  $f_e : [-1, 1] \rightarrow [0, \phi]$  for some  $\phi > 0$ . Then, for every constant  $\eta > 0$ , the expected number of steps in every implementation of the FLIP method for MAX- $k$ -CUT is  $O(\phi n^{2(2k-1)k \log(kn)+3+\eta})$ .*

**PROOF.** Let  $L$  denote the sequence of moves executed by the algorithm and let  $T := \lfloor \ell(L)/kn \rfloor$ . If  $T \geq 1$ , then by Lemma 4.20, there exists a starting configuration  $\tau_0$  and a  $k$ -cyclic block  $B$  in  $L$  such that  $B$  is  $(n/T)$ -slowly improving from  $\tau_0$  with respect to  $X$ .

For non-negative integers  $\ell'$  and  $t$ , let  $R(\ell', t)$  denote the event that there exists a  $k$ -cyclic block  $B$  in  $L$  of length  $\ell'$  and a starting partition  $\tau_0$  such that  $B$  is  $(n/t)$ -slowly improving from  $\tau_0$  with respect to edge weights  $X$ . Let

$$\zeta := \left\lceil k\phi n^{2(2k-1)k \log(kn)} \right\rceil,$$

and  $t_i := i\zeta$  for  $i \geq 2$ . We recall that  $\ell(L) \leq k^n$ , since the algorithm never cycles back to the same configuration twice. Therefore,

$$\begin{aligned} \mathbb{E}_X(T) &\leq \sum_{t=1}^{k^n} \Pr_X(T \geq t) \\ &\leq 2\zeta + \sum_{i=2}^{k^n} \sum_{t=t_i}^{t_{i+1}-1} \Pr_X(T \geq t) \\ &\leq 2\zeta + \sum_{i=2}^{k^n} \sum_{t=t_i}^{t_{i+1}-1} \Pr_X(T \geq t_i) \\ &\leq 2\zeta + \sum_{i=2}^{k^n} \sum_{t=t_i}^{t_{i+1}-1} \sum_{\ell'=1}^{kn} \Pr_X(R(\ell', t_i)). \end{aligned}$$

We will show that the second term in the RHS above is  $O(\zeta n \log n)$ .

By Lemma 4.22 and the fact that the number of  $k$ -cyclic blocks  $B$  with  $\ell(B) = \ell'$  is at most  $n^{\ell'}$ , we have that

$$\Pr_X(R(\ell', t_i)) \leq (kn)^{\ell'} \left( \frac{k\phi n}{t_i} \right)^{\frac{1}{2(2k-1)k \log(kn)} \ell'} = \left( \frac{1}{i} \right)^{\frac{1}{2(2k-1)k \log(kn)} \ell'}.$$

The last equality above is by the choice of  $\zeta$ . Hence,

$$\begin{aligned} \sum_{\ell'=1}^{kn} \Pr_X(R(\ell', t_i)) &\leq \sum_{\ell'=1}^{kn} \left( \frac{1}{i} \right)^{\frac{1}{2(2k-1)k \log(kn)} \ell'} \\ &\leq \sum_{\ell''=0}^{\infty} \sum_{\ell'''=0}^{31} \left( \frac{1}{i} \right)^{\ell'' + \frac{1}{2(2k-1)k \log(kn)} \ell'''} \\ &\leq \sum_{\ell''=1}^{\infty} 2(2k-1)k \log(kn) \left( \frac{1}{i} \right)^{\ell''} \\ &\leq 2(2k-1)k \log(kn) \left( \sum_{\ell''=0}^{\infty} \frac{1}{i^{\ell''}} - 1 \right) \\ &= \frac{2(2k-1)k \log(kn)}{i-1}. \end{aligned}$$

Therefore,

$$\begin{aligned} \sum_{i=2}^{k^n} \sum_{t=t_i}^{t_{i+1}-1} \sum_{\ell'=1}^{kn} \Pr_X(R(\ell', t_i)) &\leq \sum_{i=2}^{k^n} \sum_{t=t_i}^{t_{i+1}-1} \frac{2(2k-1)k \log(kn)}{i-1} \\ &= \sum_{i=2}^{k^n} \zeta \frac{2(2k-1)k \log(kn)}{i-1} \end{aligned}$$

$$\begin{aligned}
&= 2\zeta(2k-1)k \log(kn) O(\log(k^n)) \\
&= O(\zeta n \log n).
\end{aligned}$$

We have thus shown that  $\mathbb{E}_X(T) = O(\zeta n \log n)$  and hence, the expected number of flips performed by the algorithm is  $O(\zeta n^2 \log n) = O(\phi n^{2(2k-1)k \log(kn)+3+\eta})$  for every constant  $\eta > 0$ .  $\square$

## APPENDIX

### A PROOF OF LEMMA 3.12

We restate and prove Lemma 3.12.

**LEMMA 3.12.** *Let  $\phi > 0$  and  $X_1, \dots, X_m$  be independent random variables with density functions  $f_1, \dots, f_m : \mathbb{R} \rightarrow [0, \phi]$ . Let  $X := (x_1, \dots, x_m)^\top$  and  $\alpha_1, \dots, \alpha_k \in \mathbb{Z}^n$  be linearly independent vectors. Then for every  $\epsilon > 0$ ,*

$$\Pr_X \left[ \langle \alpha_i, X \rangle > 0 \forall i \in [k] \text{ and } \sum_{i=1}^k \langle \alpha_i, X \rangle \leq \epsilon \right] \leq \frac{(\phi\epsilon)^k}{k!}.$$

**PROOF.** Our proof closely resembles that of Lemma A.1 in Reference [7]. Let  $e_i$  denote the  $i$ th coordinate vector. We can extend  $\{\alpha_1, \dots, \alpha_k\}$  to a basis for  $\mathbb{R}^m$  by adding coordinate vectors. Without loss of generality suppose that the derived basis is  $\{\alpha_1, \dots, \alpha_k, e_{k+1}, \dots, e_m\}$ . Let

$$B := \begin{bmatrix} \alpha_1^\top \\ \vdots \\ \alpha_k^\top \\ e_{k+1}^\top \\ \vdots \\ e_m^\top \end{bmatrix}.$$

We note that  $A$  is a full-rank  $m \times m$  matrix. Let  $Y := (Y_1, \dots, Y_m)^\top := BX$  and  $A := B^{-1}$ . We note that  $Y$  is a random vector whose coordinates are possibly dependent. Let  $f, g : \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0}$  denote the joint density function of  $X$  and  $Y$ , respectively.

*Claim 8.* Let  $y_1, \dots, y_m \in \mathbb{R}$ . Then  $g(y_1, \dots, y_m) \leq \phi^k \prod_{i=k+1}^m f_i(y_i)$ .

**PROOF.** Let  $y := (y_1, \dots, y_m)$ . Then, we have  $g(y) = |\det(A)^{-1}| f(A^{-1}(y))$ . Since  $A$  is integral, we have that  $|\det(A)| \in \mathbb{Z}_+$  and hence  $|\det(A)^{-1}| \leq 1$ . Therefore,

$$\begin{aligned}
g(y) &\leq f(A^{-1}(y)) \\
&= f(A_1^{-1}(y), \dots, A_m^{-1}(y)) \\
&= f_1(A_1^{-1}(y)) \dots f_m(A_m^{-1}(y)) \\
&\quad \text{(since } X_1, \dots, X_m \text{ are independent)} \\
&\leq \phi^k f_{k+1}(A_{k+1}^{-1}(y)) \dots f_m(A_m^{-1}(y)) \\
&= \phi^k f_{k+1}(B_{k+1}(y)) \dots f_m(B_m(y)) \\
&= \phi^k f_{k+1}(y_{k+1}) \dots f_m(y_m). \quad \square
\end{aligned}$$

Let  $V := \{(y_1, \dots, y_m) \in \mathbb{R}^m : y_1, \dots, y_k > 0, \sum_{i=1}^k y_i \leq \epsilon\}$  and  $U := \{(y_1, \dots, y_k) \in \mathbb{R}^k : y_1, \dots, y_k > 0, \sum_{i=1}^k y_i \leq \epsilon\}$ . We have that

$$\begin{aligned}
\Pr_X \left[ \alpha_i^\top X > 0 \forall i \in [k], \sum_{i=1}^k \alpha_i^\top X \leq \epsilon \right] &= \Pr_Y \left[ y_i > 0 \forall i \in [k], \sum_{i=1}^k Y_i \leq \epsilon \right] \\
&= \int_V g(y_1, \dots, y_m) dy_1 \dots dy_m \\
&\leq \int_V \phi^k \left( \prod_{i=k+1}^m f_i(y_i) \right) dy_1 \dots dy_m \\
&= \phi^k \left( \int_U dy_1 \dots dy_k \right) \\
&\times \left( \int_{y_{k+1}, \dots, y_m \in \mathbb{R}} \prod_{i=k+1}^m f_i(y_i) dy_{k+1} \dots dy_m \right) \\
&\leq \phi^k \text{Vol}(U) \left( \prod_{i=k+1}^m \left( \int_{y_i \in \mathbb{R}} f_i(y_i) dy_i \right) \right) \\
&= \phi^k \text{Vol}(U) \\
&= \phi^k \frac{\epsilon^k}{k!}.
\end{aligned}$$

The last step above is because  $U$  is a  $k$ -dimensional simplex of side length  $\epsilon$ . □

## ACKNOWLEDGMENTS

We thank Alexandra Kolla for bringing this problem to our attention.

## REFERENCES

- [1] O. Angel, S. Bubeck, Y. Peres, and F. Wei. 2017. Local max-cut in smoothed polynomial time. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing (STOC'17)*. 429–437.
- [2] O. Angel and V. Tassion. 2018. Exponentially long improving sequences for MAX-CUT. Personal Communication.
- [3] A. Bhalgat, T. Chakraborty, and S. Khanna. 2010. Approximating pure Nash equilibrium in cut, party affiliation, and satisfiability games. In *Proceedings of the 11th ACM Conference on Electronic Commerce (EC'10)*. 73–82.
- [4] S. Boodaghians, R. Kulkarni, and R. Mehta. 2020. Smoothed efficient algorithms and reductions for network coordination games. In *Proceedings of the 11th Innovations in Theoretical Computer Science Conference (ITCS'20)*. 73:1–73:15.
- [5] R. Elsässer and T. Tscheuschner. 2011. Settling the complexity of local max-cut (almost) completely. In *Proceedings of the 38th International Colloquium Conference on Automata, Languages and Programming (ICALP'11)*. 171–182.
- [6] M. Etscheid and H. Röglin. 2015. Smoothed analysis of the squared euclidean maximum-cut problem. In *Proceedings of the European Symposium on Algorithms (ESA'15)*. 509–520.
- [7] M. Etscheid and H. Röglin. 2017. Smoothed analysis of local search for the maximum-cut problem. *ACM Trans. Algor.* 13, 2, Article 25 (2017), 12 pages.
- [8] A. Fabrikant, C. Papadimitriou, and K. Talwar. 2004. The complexity of pure Nash equilibria. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC'04)*. 604–612.
- [9] A. Frieze and M. Jerrum. 1997. Improved approximation algorithms for MAX $k$ -CUT and MAX BISECTION. *Algorithmica* 18, 1 (May 1997), 67–81.
- [10] D. Johnson, C. Papadimitriou, and M. Yannakakis. 1988. How easy is local search? *J. Comput. Syst. Sci.* 37, 1 (1988), 79–100.
- [11] J. Kleinberg and É. Tardos. 2006. *Algorithm Design*. Addison-Wesley.
- [12] A. Schäffer and M. Yannakakis. 1991. Simple local search problems that are hard to solve. *SIAM J. Comput.* 20, 1 (1991), 56–87.
- [13] D. Spielman and S.-H. Teng. 2004. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *J. ACM* 51, 3 (2004), 385–463.

Received September 2020; revised March 2021; accepted March 2021

ACM Transactions on Algorithms, Vol. 17, No. 3, Article 19. Publication date: July 2021.