The Karger-Stein Algorithm Is Optimal for k-Cut

Anupam Gupta
anupamg@cs.cmu.edu
Carnegie Mellon University
Pittsburgh, PA, USA

Euiwoong Lee
euiwoong@cims.nyu.edu
NYU
New York, NY, USA

Jason Li
jmli@cs.cmu.edu
Carnegie Mellon University
Pittsburgh, PA, USA

ABSTRACT

In the k-cut problem, we are given an edge-weighted graph and want to find the least-weight set of edges whose deletion breaks the graph into k connected components. Algorithms due to Karger-Stein and Thorup showed how to find such a minimum k-cut in time approximately $O(n^{2k-2})$. The best lower bounds come from conjectures about the solvability of the k-clique problem and a reduction from k-clique to k-cut, and show that solving k-cut is likely to require time $\Omega(n^k)$. Our recent results have given special-purpose algorithms that solve the problem in time $n^{1.98k+O(1)}$, and ones that have better performance for special classes of graphs (e.g., for small integer weights).

In this work, we resolve the problem for general graphs, by showing that for any fixed $k \geq 2$, the Karger-Stein algorithm outputs any fixed minimum k-cut with probability at least $\widehat{O}(n^{-k})$, where $\widehat{O}(\cdot)$ hides a $2^{O(\ln \ln n)^2}$ factor. This also gives an extremal bound of $\widehat{O}(n^k)$ on the number of minimum k-cuts in an n-vertex graph and an algorithm to compute a minimum k-cut in similar runtime. Both are tight up to $\widehat{O}(1)$ factors.

The first main ingredient in our result is a fine-grained analysis of how the graph shrinks—and how the average degree evolves—under the Karger-Stein process. The second ingredient is an extremal result bounding the number of cuts of size at most $(2-\delta)OPT/k$, using the Sunflower lemma.

CCS CONCEPTS

• **Theory of computation → Graph algorithms analysis**.

KEYWORDS

Randomized algorithms, Minimum Cut, Karger-Stein Algorithms, Sunflower Lemma

ACM Reference Format:

Anupam Gupta, Euiwoong Lee, and Jason Li. 2020. The Karger-Stein Algorithm Is Optimal for k-Cut. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC '20), June 22–26, 2020, Chicago, IL, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3357713.3384285

STOC '20, June 22–26, 2020, Chicago, IL, USA

ACM ISBN 978-1-4503-6979-4/20/06…$15.00
https://doi.org/10.1145/3357713.3384285

1 INTRODUCTION

We consider the k-Cut problem: given an edge-weighted graph $G = (V, E, w)$ and an integer k, delete a minimum-weight set of edges so that G has at least k connected components. This problem generalizes the global min-cut problem, where the goal is to break the graph into $k = 2$ pieces. It was unclear that the problem admitted a polynomial-time algorithm for fixed values of k, until the work of Goldschmidt and Hochbaum, who gave a runtime of $O(n^{(1/2-o(1))k^2})$ [GH94]. (Here and subsequently, the $o(1)$ in the exponent indicates a quantity that goes to 0 as k increases.) The randomized minimum-cut algorithm of Karger and Stein [KS96], based on random edge contractions, can be used to solve k-Cut in $\tilde{O}(n^{2(k-1)})$ time. For deterministic algorithms, there have been improvements to the Goldschmidt and Hochbaum result [KYN07, Tho08, CQX18]: notably, the tree-packing result of Thorup [Tho08] was sped up by Chekuri et al. [CQX18] to run in $O(mn^{2k-3})$ time. Hence, until recently, randomized and deterministic algorithms using very different approaches achieved $O(n^{(2-o(1))k})$ bounds for the problem.

For hardness, we reduce Max-Weight $(k-1)$-Clique to k-Cut. It is conjectured that solving Max-Weight $(k-1)$-Clique requires $\tilde{\Omega}(n^{(1-o(1))k})$ time when weights are integers in the range $[1, \Omega(n^k)]$, and $\tilde{\Omega}(n^{(\omega/3)k})$ time for unit weights; here ω is the matrix multiplication constant. Hence, these runtime lower bounds also extend to the k-Cut, suggesting that $\Omega(n^k)$ may be the optimal runtime for general weighted k-cut instances.

There has been recent progress on this problem, showing the following results:

(1) We showed an $O(n^{(1.98+o(1))k})$-time algorithm for general k-Cut [GLL19]. This was based on giving an extremal bound on the maximum number of "small" cuts in the graph, and then using a bounded-depth search approach to guess the small cuts within the optimal k-cut and make progress. This was a proof-of-concept result, showing that the bound of $n^{(2-o(1))k}$ was not the right bound, but it does not seem feasible to improve that approach to exponents considerably below $2k$.

(2) For graphs with polynomial integer weights, we showed an algorithm to solve the problem in time approximately $k^{O(k)}\, n^{(2\omega/3+o(1))k}$ [GLL18]. And for unweighted graphs we showed how to get the $k^{O(k)} n^{(1+o(1)k}$ runtime [Li19]. Both these approaches were based on obtaining a spanning tree cut by a minimum k-cut in a small number of edges, and using involved dynamic programming methods on the tree to efficiently compute the edges and find the k-cut. The former relied on matrix multiplication ideas, and the latter on the Kawarabayashi-Thorup graph decomposition, both of which are intrinsically tied to graphs with small edge-weights.

In this paper, we show that the "right" algorithm, the original Karger-Stein algorithm, achieves the "right" bound for general graphs. Our main result is the following.

Theorem 1 (Main). *Given a graph G and a parameter $k \geq 2$, the Karger-Stein algorithm outputs any fixed minimum k-cut in G with probability at least $n^{-k} \cdot (k \ln n)^{-O(k^2 \ln \ln n)}$.*

For any fixed constant $k \geq 2$, the above bound becomes $n^k \cdot 2^{O(\ln \ln n)^2}$. This immediately implies the following two corollaries, where $\widehat{O}(\cdot)$ hides a quasi-logarithmic factor $2^{O(\ln \ln n)^2}$:

Corollary 2 (Number of Minimum k-cuts)**.** *For any fixed $k \geq 2$, the number of minimum-weight k-cuts in a graph is at most $\widehat{O}(n^k)$.*

This improves on the previous best bound of $n^{(1.98+o(1))k}$ [GLL19]. It is also almost tight because the cycle on n vertices has $\Omega(n^k)$ minimum k-cuts.

Corollary 3 (Faster Algorithm to Find a Minimum k-cut)**.** *For any fixed $k \geq 2$, there is a randomized algorithm that computes a minimum k-cut of a graph with high probability in time $\widehat{O}(n^k)$.*

This improves the running time $n^{(1.98+o(1))k}$ for the general weighted case [GLL19] and even the running time $n^{(1+o(1))k}$ for the unweighted case [Li19], where the extra $n^{o(k)}$ term is still at least polynomial for fixed k. It is also almost tight under the hypothesis that Max-Weight $(k-1)$-Clique requires $\tilde{\Omega}(n^{(1-o(1))k})$ time.

1.1 Our Techniques

Let us first recall the Karger-Stein algorithm:

Algorithm 1 Karger-Stein Algorithm

procedure Karger-Stein($G = (V, E, w)$, $k \in \mathbb{N}$) ▷ Compute a minimum k-cut of G
  while $|V| > k$ **do**
    Sample edge $e \in E$ with prob. proportional to $w(e)$.
    Contract two vertices in e and remove self-loops.
  end while
  Return the k-cut of G by expanding vertices of V.
end procedure

In the spirit of [GLL19], our proof consists of two main parts: (i) a new algorithmic analysis (this time for the Karger-Stein algorithm), and (ii) a statement on the extremal number of "small" cuts in a graph. In order to motivate the new analysis of Karger-Stein, let us first state a crude version of our extremal result. Define λ_k as the minimum k-cut value of the graph. Think of $\overline{\lambda}_k := \lambda_k/k$ as the average contribution of each of the k components to the k-cut. Loosely speaking, the extremal bound says the following:

> ($\star$) There are at most a *linear* number of 2-cuts of a graph of size at most $1.99\overline{\lambda}_k$.

To develop some intuition for this claim, we make two observations about the cycle and clique graphs, two graphs where the number of minimum k-cuts is indeed $\Omega(n^k)$. Firstly, in the n-cycle, $\overline{\lambda}_k = 1 = \lambda/2$ for any value of k, and since $1.99\overline{\lambda}_k = 1.99 < 2 = \lambda$, so there are *no* cuts in the graph with size at most $1.99\overline{\lambda}_k$, hence ($\star$) holds. However, it breaks if $1.99\overline{\lambda}_k$ is replaced by $2\overline{\lambda}_k$, since there are $\binom{n}{2}$ many cuts of size $2\overline{\lambda}_k = 2 = \lambda$. Secondly, for the n-clique we have $\lambda_k \approx k(n-1)$, since the minimum k-cut chops off $k-1$ singleton vertices. (We assume $k \ll n$, and ignore the $\binom{k}{2}$ double-counted edges for simplicity.) We have $\overline{\lambda}_k \approx n - 1 = \lambda$ instead for the n-clique, and there are exactly n cuts of size at most $1.99\overline{\lambda}_k$ (the singletons), so our bound ($\star$) holds. And again, ($\star$) fails when $1.99\overline{\lambda}_k$ replaced by 2. Therefore, in both the cycle and the clique, the bound $1.99\overline{\lambda}_k$ is almost the best possible. Moreover, the linear bound in the number of cuts is also optimal in the clique. In general, it is instructive to consider the cycle and clique as two opposite ends of the spectrum in the context of graph cuts, since one graph has n minimum cuts and the other has $\binom{n}{2}$.

1.1.1 Algorithmic Analysis. To analyze Karger-Stein, we adopt an *exponential clock* view of the process: fix an infinitesimally small parameter δ, and on each timestep of length δ, sample each edge with an independent probability $\delta/\overline{\lambda}_k$ and contract it. If δ is small enough, then we can disregard the event that more than one edge is sampled on a single timestep. This perspective has a distinct advantage over the classical random contraction procedure that contracts one edge at a time: we can analyze whether each edge is contracted *independently*. At the same time, we reemphasize it is just another view of the same process; conditioned on the number of vertices n' in the remaining graph, the outcome of the exponential clock procedure has exactly the same distribution as the standard Karger-Stein procedure that iteratively contracted $n - n'$ edges.

Let us first translate the classical Karger-Stein analysis in the exponential clock setting. Suppose that we run the process for $2 \ln n$ units of time (that is, $\frac{2 \ln n}{\delta}$ timesteps of length δ each). The probability that a fixed minimum k-cut remains at the end (i.e., has no edge contracted) is roughly

$$\left(1 - \lambda_k \cdot \frac{\delta}{\overline{\lambda}_k}\right)^{(2 \ln n)/\delta} \approx \exp\left(-\lambda_k \cdot \frac{\delta}{\overline{\lambda}_k} \cdot \frac{2 \ln n}{\delta}\right) = n^{-2k}.$$

How many vertices are there remaining after $2 \ln n$ units of time? On each δ-timestep, consider the graph before any edges are contracted on this timestep. If there are currently $r \geq k-1$ vertices, then the sum of the degrees of the vertices must be at least $r\overline{\lambda}_k$; otherwise, we can cut out $k-1$ random singletons to obtain a k-cut of size at most $(k-1) \cdot \overline{\lambda}_k < \lambda_k$ (which corresponds to a k-cut of the same size in the original graph once we "uncontract" each edge), contradicting the definition of λ_k as the minimum k-cut. Therefore, the graph has at least $r\overline{\lambda}_k/2$ edges, which means that we contract at least $r\overline{\lambda}_k/2 \cdot \delta/\overline{\lambda}_k = r\delta/2$ edges in expectation. We now claim that as $\delta \to 0$, this is essentially equivalent to contracting at least $r\delta/2$ *vertices* in expectation, since we should not contract more than one edge at any timestep. That is, we contract a $\delta/2$ *fraction* of the vertices per timestep, in expectation. Therefore, after $2 \ln n$ units of time (which is $\frac{2 \ln n}{\delta}$ timesteps), the expected number of vertices remaining is at most

$$n \cdot \left(1 - \frac{\delta}{2}\right)^{(2 \ln n)/\delta} \approx n \cdot \exp\left(-\frac{\delta}{2} \cdot \frac{2 \ln n}{\delta}\right) = n \cdot \frac{1}{n} = 1.$$

Recall that the Karger-Stein algorithm terminates when there are k vertices left. Therefore, informally, we should expect to be done before time $2\ln n$. This argument is not rigorous, since we cannot take a naïve union bound over the two successful events (namely, that no edge in the fixed k-cut is contracted, and there are $k-1$ vertices remaining at the end). We elaborate on how to handle this issue later on.

At any given step of the Karger-Stein process, the bound of $r\overline{\lambda}_k/2$ edges can be tight in the worse case. So instead of improving the analysis in a worst-case scenario, the main insight to our improvement is a more average-case improvement. At a high level, as the process continues, we expect more and more vertices in the contracted graph to have degree *much higher* than $\overline{\lambda}_k$. In fact, we show that the fraction of vertices with degree at most $1.99\overline{\lambda}_k$ is expected to shrink significantly throughout the process. Consequently, the total sum of degrees becomes much larger than $r\overline{\lambda}_k$, from which we obtain an improvement.

How do we obtain such a guarantee? Observe that every vertex of degree at most $1.99\overline{\lambda}_k$ in the graph at some intermediate stage of the process corresponds to a (2-)cut in the original graph of size at most $1.99\overline{\lambda}_k$. Recall that by our extremal bound, we start off with only $O_k(n)$ many such cuts. Some of these cuts can have size less than $\overline{\lambda}_k$, but we show that there cannot be too many: at most 2^{k-1} of them. The more interesting case is cuts of size in the range $[\overline{\lambda}_k, 1.99\overline{\lambda}_k]$: since each of these cuts has size at least $\overline{\lambda}_k$, the probability that we contract an edge in a fixed cut is at least $\overline{\lambda}_k \cdot \delta/\overline{\lambda}_k = \delta$. This means that after just $\ln n$ units of time (and not $2\ln n$), the cut remains intact only with probability $(1-\delta)^{(\ln n)/\delta} \approx 1/n$. Taking an expectation over all $O_k(n)$ cuts, we expect only $O_k(1)$ of them to remain after $\ln n$ time, which means we expect only $O_k(1)$ many vertices of degree at most $1.99\overline{\lambda}_k$ in the contracted graph after $\ln n$ time has passed. This analysis works for any time t: we expect only $O_k(n)\cdot(1-\delta)^{t/\delta}$ vertices of degree at most $1.99\overline{\lambda}_k$ in the contracted graph after time t.

This upper bound on the number of small-degree vertices lower bounds the number of edges in the graph, which in turn governs the rate at which the graph shrinks throughout the process. To obtain the optimal bounds, we model the expected rate of decrease as a differential equation. In expectation, we find that by time $\ln n$ (not $2\ln n$), we expect only $\widehat{O}_k(1)$ vertices remaining. This is perfect, since a fixed minimum k-cut survives with probability n^{-k} by this time! To finish off the analysis from $\widehat{O}_k(1)$ vertices down to $(k-1)$ vertices, we use the regular Karger-Stein analysis, picking up an extra factor of $\left(\widehat{O}_k(1)\right)^k = \widehat{O}_k(1)$.

Lastly, the issue of the $\widehat{O}_k(1)$ vertices bound holding only in expectation requires some technical work to handle. In essence, we strengthen the expectation statement to one with high probability. We then union-bound over the event of $\widehat{O}_k(1)$ vertices remaining and the event that the minimum k-cut survives—which only holds with probability n^{-k}. This requires us to use concentration bounds combined with a recursive approach; the details appear in §2.2 and §2.3.

1.1.2 Extremal Result. Recall our target extremal statement (★): there are $O_k(n)$ many cuts of a graph of size at most $1.99\overline{\lambda}_k$. Suppose for contradiction that there are $\omega_k(n)$ such cuts, and assume for simplicity that k is even. Our goal is to select $k/2$ of these cuts that "cross in many ways": namely, there are at least k nonempty regions in their Venn diagram (see Figure 1 left). This gives a k-cut with total cost $k/2 \cdot 1.99\overline{\lambda}_k < \lambda_k$, contradicting the definition of λ_k as the minimum k-cut.

To find such a collection of crossing cuts, one approach is to treat each cut as a subset of vertices (the vertices on one side of the cut), and tackle the problem from a purely extremal set-theory perspective, ignoring the underlying structure of the graph. The statement becomes: *given a family of $\omega_k(n)$ many distinct subsets of $[n]$, there exist some $k/2$ subsets whose Venn diagram has at least k nonempty atoms.* In [GLL19], we tackled a similar problem from this point of view. However, for our present problem, the corresponding extremal set theory statement is too good to be true. The set system $([n], \binom{[n]}{2})$ has $\binom{n}{2}$ sets, but no $k/2 = 3$ sets can possibly form $k = 6$ regions. Hence, we need the additional structure of cuts in a graph to formulate an extremal set theoretic statement that holds.

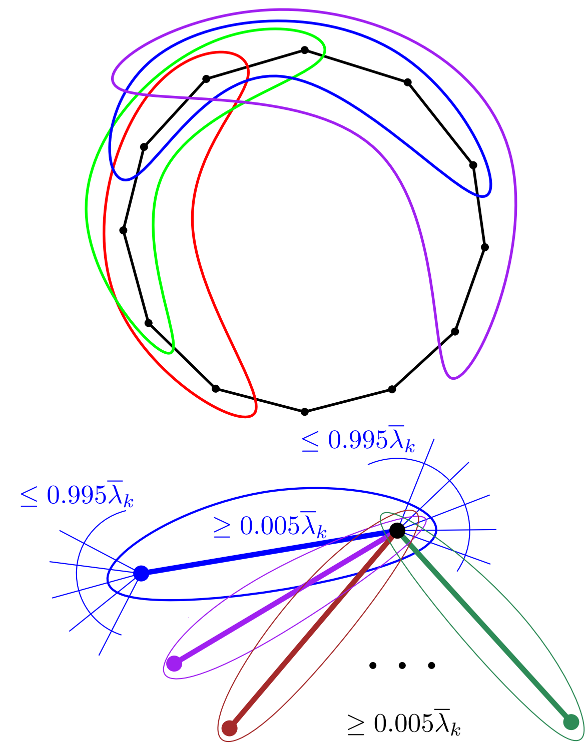

Figure 1: Top: For a contradiction, suppose all $\binom{n}{2}$ cuts of the cycle have size at most $1.99\overline{\lambda}_k$. Then, we select $k/2 = 4$ many such cuts as shown. Their Venn diagram has 8 nonempty atoms and form an 8-cut with cost $4 \cdot 1.99\overline{\lambda}_k = 7.92\overline{\lambda}_k < 8\overline{\lambda}_k = \lambda_k$, contradicting the definition of λ_k as the minimum k-cut. Bottom: A 401-sunflower with nonempty core, with the core and all petals contracted to single vertices, each of degree at least $\overline{\lambda}_k$. Each bolded edge must have total weight at least $0.005\overline{\lambda}_k$ for the corresponding cut to have size at most $1.99\overline{\lambda}_k$. However, the 400 many bolded edges excluding the blue one give total weight at least $400\cdot 0.005\overline{\lambda}_k = 2\overline{\lambda}_k$, and each of them crosses the blue cut, contradicting the assumption that the blue cut has size at most $1.99\overline{\lambda}_k$.

Our key observation is that the cut structure of the graph forbids large *sunflowers with nonempty core* in the corresponding set family. To see why, consider a 401-sunflower of sets $S_1, S_2, \dots, S_{401}$ with nonempty core C, and suppose in addition that the core C and each petal $S_i \setminus C$ is a cut of size at least $\overline{\lambda}_k$. (Handling cuts of size less than $\overline{\lambda}_k$ is a technical detail, so we omit it here.) For simplicity, consider contracting the core C and petals $S_i \setminus C$ into single vertices c and p_i, respectively. For each $i \in [401]$, the vertices c and p_i each have degree at least $\overline{\lambda}_k$, and yet the cut $\{c, p_i\}$ has size at most $1.99\overline{\lambda}_k$; a simple calculation shows that there must be at least $0.005\overline{\lambda}_k$ edges between c and p_i. Equivalently, in the weighted case, the edge (c, p_i) has weight at least $0.005\overline{\lambda}_k$ (see Figure 1 right). Now observe that the edges $(c, p_2), (c, p_3), \dots, (c, p_{401})$ all cross the cut $\{c, p_1\}$, and together, they have total weight $400 \cdot 0.005\overline{\lambda}_k = 2\overline{\lambda}_k$. Hence, the cut $\{c, p_1\}$ must have weight at least $2\overline{\lambda}_k$, contradicting the assumption that it has size at most $1.99\overline{\lambda}_k$.

With this insight in mind, our modified extremal set theory statement is as follows (when k is even): given a family of $\omega_k(n)$ many distinct subsets of $[n]$ that *do not contain a* 401*-sunflower with nonempty core*, there exist some $k/2$ many subsets whose Venn diagram has at least k nonempty atoms. This statement turns out to be true, and we provide a clean inductive argument that uses the Sunflower Lemma as a base case. (Our actual extremal statement is slightly different to handle the cuts of size less than $\overline{\lambda}_k$ as well as odd k.)

1.2 Preliminaries

A weighted graph is denoted by $G = (V, E, w)$ where $w : E \to \mathbb{Q}^+$ gives positive rational weights on edges. Let λ_k be the weight of a minimum k-cut, and $\overline{\lambda}_k := \lambda_k/k$. We use different logarithms to naturally present different parts of our analysis. We use ln to denote $\log_e$ and lg to denote $\log_2$.

2 RANDOM GRAPH PROCESS

In this section, we analyze the exponential clock viewpoint of the Karger-Stein algorithm to prove our main result Theorem 1, assuming the bound on the number of small cuts (Theorem 10) proved in Section 3.

For the sake of exposition, throughout this section, we assume G is an unweighted multigraph, where the number of edges between a pair (u, v) is proportional to its weight. Note that while it changes the number of edges m and $\overline{\lambda}_k$, the resulting exponential clock procedure is still exactly equivalent to the standard Karger-Stein procedure in the original weighted graph, and the final bounds in the main lemmas (Lemma 4 and Lemma 6) only involve n, k, t.

2.1 Expectation

In this section, we bound the number of vertices (in expectation) at some point in the random contraction process:

Lemma 4. *Suppose G has at most βn many cuts with weight in the range $[\overline{\lambda}_k, \gamma\overline{\lambda}_k)$ for some constant $1 \le \gamma < 2$. Fix a parameter $t \ge 0$, and suppose we contract every edge in G with independent probability $1 - e^{-t/\overline{\lambda}_k}$. Then, the expected number of vertices in the contracted graph is at most $O(\frac{\beta+1}{2-\gamma})ne^{-(\gamma/2)t} + \frac{\gamma}{2}(k-1)t$.*

Consider the following *exponential clock process*: let each edge e of the graph independently sample a random variable $x(e)$ from an exponential distribution with mean $\overline{\lambda}_k$, which has c.d.f. $1 - e^{-t/\overline{\lambda}_k}$ at value t. We say that the edge is *sampled at time* $x(e)$. Observe that for any $t \ge 0$, every edge is sampled by time t with probability exactly $1 - e^{-t/\overline{\lambda}_k}$, so this process models exactly the one in the lemma.

For a given $t \ge 0$ and $\delta > 0$, the probability that an edge e is sampled before time $t + \delta$, *given* that it is sampled after time t, is

$$\frac{\Pr[t \le x(e) \le t + \delta]}{\Pr[t \le x(e)]} = \Pr[x(e) \le \delta] = 1 - e^{-\delta/\overline{\lambda}_k} = \frac{\delta}{\overline{\lambda}_k} - O_{\overline{\lambda}_k}(\delta^2),$$

where the first equality uses the *memoryless* property of exponential random variables, and the $O_{\overline{\lambda}_k}(\cdot)$ hides the dependence on $\overline{\lambda}_k$. Therefore, at the loss of the $O_{\overline{\lambda}_k}(\delta^2)$ factor (which we will later show to be negligible), we can imagine the "discretized" process at a small timestep δ: the times t are now integer multiples of δ, and each edge e is sampled at the (discrete) time t if $t \le x(e) < t + \delta$. Again, the probability that an edge is sampled at (discrete) time $t + \delta$, given that it is sampled after time t, is $\delta/\overline{\lambda}_k + O_{\overline{\lambda}_k}(\delta^2) \approx \delta/\overline{\lambda}_k$.

Consider the graph at a given (discrete) time t, where we have contracted all edges sampled before time t (but not those sampled at time t). Suppose that there are $r = r(t)$ vertices in the contracted graph and $s = s(t)$ of them have degree in the range $[\overline{\lambda}_k, \gamma\overline{\lambda}_k)$. Note that at all times, at most $k - 1$ vertices have degree $< \overline{\lambda}_k$, since otherwise, we could take $(k-1)$ of those vertices (leaving at least one vertex remaining) and form a k-cut of weight less than $(k-1)\overline{\lambda}_k < \lambda_k$. Therefore, there are at least $r - s - (k-1)$ vertices with degree greater than $\gamma\overline{\lambda}_k$, so the number of edges is at least

$$\begin{aligned}\frac{s \cdot \overline{\lambda}_k + (r - s - (k-1)) \cdot \gamma\overline{\lambda}_k}{2} &= \frac{\gamma r - \gamma s + s - \gamma(k-1)}{2}\overline{\lambda}_k \\ &= \left(\frac{\gamma}{2}r - \frac{\gamma-1}{2}s - \frac{\gamma}{2}(k-1)\right)\overline{\lambda}_k.\end{aligned}$$

For now, fix $\delta > 0$, where t is now not necessarily an integer multiple of δ, and consider the time interval $[t, t + \delta)$, where we contract all edges e with $x(e) \in [t, t + \delta)$. Each edge in the current contracted graph is contracted with probability $\delta/\overline{\lambda}_k - O_{\overline{\lambda}_k}(\delta^2)$ in this time interval, so we expect at least

$$\begin{aligned}&\left(\frac{\gamma}{2}r - \frac{\gamma-1}{2}s - \frac{\gamma}{2}(k-1)\right)\overline{\lambda}_k \cdot \left(\frac{\delta}{\overline{\lambda}_k} - O_{\overline{\lambda}_k}(\delta^2)\right) \\ &= \left(\frac{\gamma}{2}r - \frac{\gamma-1}{2}s - \frac{\gamma}{2}(k-1)\right)\delta - O_{\overline{\lambda}_k, m, n}(\delta^2) \qquad (1)\end{aligned}$$

edges to be contracted, where $O_{\overline{\lambda}_k, m, n}(\cdot)$ hides dependence on $\overline{\lambda}_k, m, n$ (note that $r, s \le n$). Ideally, we now want to argue that every edge that is contracted in this interval reduces the number of remaining vertices by 1. In general, this is not true if we contract a subset of edges that contain a cycle. However, if δ is small enough (say, $\delta \ll 1/\mathrm{poly}(n)$), then in most cases, there is at most one edge contracted at all, in which case our desired argument holds.

More formally, let B_t be the (bad) event that more than one edge is contracted in the time interval $[t, t + \delta)$. Then, by a union bound over all pairs of edges, we have

$$\Pr[B_t] \le \binom{m}{2} \cdot \left(\frac{\delta}{\overline{\lambda}_k} - O_{\overline{\lambda}_k}(\delta^2)\right)^2 = O_{\overline{\lambda}_k,m,n}(\delta^2).$$

If the event B_t holds, we will apply the trivial bound $\mathbb{E}[r(t+\delta)] \le n$, and otherwise, we will use (1). We obtain

$$\begin{aligned}\mathbb{E}[r(t+\delta)] &\le (1 - \Pr[B_t]) \cdot \Big(r - \Big(\frac{\gamma}{2}r - \frac{\gamma-1}{2}s - \frac{\gamma}{2}(k-1)\Big)\delta \\ &\qquad + O_{\overline{\lambda}_k,m,n}(\delta^2)\Big) + \Pr[B_t] \cdot n \\ &\le \Big(r - \Big(\frac{\gamma}{2}r - \frac{\gamma-1}{2}s - \frac{\gamma}{2}(k-1)\Big)\delta \\ &\qquad + O_{\overline{\lambda}_k,m,n}(\delta^2)\Big) + O_{\overline{\lambda}_k,m,n}(\delta^2) \cdot n \\ &= r - \Big(\frac{\gamma}{2}r - \frac{\gamma-1}{2}s - \frac{\gamma}{2}(k-1)\Big)\delta + O_{\overline{\lambda}_k,m,n}(\delta^2).\end{aligned}$$

Taking the expectation at time t and using linearity of expectation, we obtain

$$\begin{aligned}\mathbb{E}[r(t+\delta)] \le \mathbb{E}[r(t)] - \Big(\frac{\gamma}{2}\mathbb{E}[r(t)] - \frac{\gamma-1}{2}\mathbb{E}[s(t)] - \frac{\gamma}{2}(k-1)\Big)\delta \\ + O_{\overline{\lambda}_k,m,n}(\delta^2).\end{aligned}$$

We now bound $s = s(t)$ in terms of t. Every vertex whose degree is in the range $[\overline{\lambda}_k, \gamma\overline{\lambda}_k)$ must correspond to a cut in G with weight in $[\overline{\lambda}_k, \gamma\overline{\lambda}_k)$, and by assumption, there are at most βn of them for some fixed constant $\beta > 0$. The probability that a cut of size $c \ge \overline{\lambda}_k$ has all its edges remaining up to time t is $e^{-ct/\overline{\lambda}_k} \le e^{-t}$, so we expect at most $e^{-t} \cdot \beta n$ of these cuts to survive by time t. Therefore, $\mathbb{E}[s(t)] \le e^{-t}\beta n$, and

$$\begin{aligned}\mathbb{E}[r(t+\delta)] \le \mathbb{E}[r(t)] - \Big(\frac{\gamma}{2}\mathbb{E}[r(t)] - \frac{\gamma-1}{2}e^{-t}\beta n - \frac{\gamma}{2}(k-1)\Big)\delta \\ + O_{\overline{\lambda}_k,m,n}(\delta^2).\end{aligned}$$

Subtracting $\frac{\gamma}{2}(k-1)(t+\delta)$ from both sides, we obtain

$$\begin{aligned}\mathbb{E}\Big[r(t+\delta) - \frac{\gamma}{2}(k-1)(t+\delta)\Big] \le \mathbb{E}\Big[r(t) - \frac{\gamma}{2}(k-1)t\Big] - \\ \Big(\frac{\gamma}{2}\mathbb{E}[r(t)] - \frac{\gamma-1}{2}e^{-t}\beta n\Big)\delta + O_{\overline{\lambda}_k,m,n}(\delta^2).\end{aligned}$$

We now solve for $\mathbb{E}[r(t)]$. Define $f(t) := \mathbb{E}[r(t) - \frac{\gamma}{2}(k-1)t]$, so that

$$\begin{aligned}f(t+\delta) - f(t) &\le -\Big(\frac{\gamma}{2}\Big[f(t) + \frac{\gamma}{2}(k-1)t\Big] - \frac{\gamma-1}{2}e^{-t}\beta n\Big)\delta \\ &\qquad + O_{\overline{\lambda}_k,m,n}(\delta^2) \\ &\le -\Big(\frac{\gamma}{2}f(t) - \frac{\gamma-1}{2}e^{-t}\beta n\Big)\delta + O_{\overline{\lambda}_k,m,n}(\delta^2).\end{aligned}$$

Taking $\delta \longrightarrow 0$, we obtain the differential equation

$$f'(t) = \lim_{\delta\to 0}\frac{f(t+\delta) - f(t)}{\delta} \le -\frac{\gamma}{2}f(t) + \frac{\gamma-1}{2}e^{-t}\beta n.$$

Set $B := \frac{\gamma-1}{2}\beta$, so we instead have

$$f'(t) \le -\frac{\gamma}{2}f(t) + Be^{-t}n. \tag{2}$$

Observe that if we had $f'(t) = -\frac{\gamma}{2}f(t)$ instead, then that would solve to $f(t) \le e^{-(\gamma/2)t}n$, but there's the additional $Be^{-t}n$ term to deal with. However, e^{-t} drops much faster than $e^{-(\gamma/2)t}$ (since $\gamma < 2$ by assumption), so intuitively, the $Be^{-t}n$ factor doesn't affect us asymptotically. We now formalize our intuition.

To upper bound $f(t)$, we will solve the differential equation (2) where we pretend the inequality in (2) is actually an equality. More formally, define

$$A := \frac{B}{B + 1 - \gamma/2},$$

which satisfies $A < 1$ since $\gamma < 2$, and define

$$\tilde{f}(t) := \frac{1}{1-A}n(e^{-(\gamma/2)t} - Ae^{-t}).$$

The following is a simple exercise in differential equations which we defer to the appendix.

Claim 5. *The function $\tilde{f}(t)$ satisfies $\tilde{f}(0) = f(0)$ and*

$$\tilde{f}'(t) = -\frac{\gamma}{2}\tilde{f}(t) + Be^{-t}n,$$

which is the differential equation (2) with equality (where f is replaced by $\tilde{f}$). Moreover, $\tilde{f}(t) \ge f(t)$ for all $t \ge 0$.

Following Claim 5, we have

$$\begin{aligned}\mathbb{E}\Big[r(t) - \frac{\gamma}{2}(k-1)t\Big] = f(t) \le \tilde{f}(t) &\le \frac{1}{1-A}n(e^{-(\gamma/2)t} - Ae^{-t}) \\ &\le \frac{1}{1-A}ne^{-(\gamma/2)t} \\ &= \frac{B + 1 - \gamma/2}{1 - \gamma/2}ne^{-(\gamma/2)t} \\ &= \frac{\frac{\gamma-1}{2}\beta + 1 - \gamma/2}{1 - \gamma/2}ne^{-(\gamma/2)t} \\ &= \Big(\frac{\gamma-1}{2-\gamma}\beta + 1\Big)ne^{-(\gamma/2)t} \\ &= O\Big(\frac{\beta+1}{2-\gamma}\Big)ne^{-(\gamma/2)t}.\end{aligned}$$

Adding $\frac{\gamma}{2}(k-1)t$ to each side finishes the proof of Lemma 4.

2.2 Concentration

In this section, we prove that for any graph with bounded number of edges, if we sample each edge independently with probability $p = 1 - e^{-t/\overline{\lambda}_k}$, the number of connected components is at most $\widetilde{O}(\sqrt{n})$ plus the expected value with high probability. It will be subsequently used in the recursive analysis in Section 2.3.

Lemma 6. *Let $\alpha \ge 1$, $t \ge \Omega(1)$, and $N \ge n$ be parameters. Let G be a graph with at most $\alpha\overline{\lambda}_k n$ edges. Suppose we sample every edge in G with independent probability $1 - e^{-t/\overline{\lambda}_k}$; let the random variable f denote the number of connected components in the sampled graph. Then, with probability at least $1 - N^{-2k}$, we have $f \le \mathbb{E}[f] + O(k \ln N\sqrt{\alpha t n})$.*

Proof. Let $e_1, \dots, e_m$ be the edges of G, arbitrarily ordered. For each $i \in [m]$, let $X_i \in \{0, 1\}$ be the random variable indicating that e_i is sampled. Then each X_i is independent and $\Pr[X_i = 1] = p$

where $p = 1 - e^{-t/\overline{\lambda}_k}$. Let $f(X_1, \ldots, X_m)$ be the number of components of the graph whose edge set is $\{e_i : X_i = 1\}$. For each $i \in [m]$, let

$$Y_i := \mathbb{E}[f(X_1, \ldots, X_m)|X_1, \ldots, X_i],$$
$$Z_i := Y_i - Y_{i-1},$$
$$W_i := \sum_{j=1}^{i} \mathbb{E}[Z_j^2|X_1, \ldots, X_{j-1}].$$

Together with $Y_0 = \mathbb{E}[f]$, the sequence $\{Y_0, \ldots, Y_m\}$ forms a Doob martingale.

Since the existence of one edge changes the number of connected components by at most 1, $|Z_i| \le 1$ always for every $i \in [m]$. For every $j \in [m]$ and $X_1, \ldots, X_{j-1}$ (which determines Y_{j-1}), let $y_b := \mathbb{E}[f|X_1, \ldots, X_{j-1}, X_j = b]$ for $b \in \{0, 1\}$. By the same argument, $|y_0 - y_1| \le 1$, $y_1 \le Y_{i-1} \le y_0$, and $Y_{i-1} := py_1 + (1-p)y_0$, so that

$$\mathbb{E}[Z_j^2|X_1, \ldots, X_{j-1}] = p\big((1-p)(y_1 - y_0)\big)^2 + (1-p)\big(p(y_0 - y_1)\big)^2 \le p(1-p).$$

In particular, $W_i \le pi$ for every $i \in [m]$ with probability 1. We use the following concentration inequality for martingales, due to Freedman.

Theorem 7. *[Fre75] Let $\{Y_0, \ldots, Y_m\}$ be a martingale with associated differences $Z_i := Y_i - Y_{i-1}$, and*

$$W_i := \sum_{j=1}^{i} \mathbb{E}[Z_i^2 \mid Y_1, \ldots, Y_{k-1}],$$

such that with probability 1, we have $|Z_i| \le R$ for every i and $W_m \le \sigma^2$. Then for all $s \ge 0$,

$$\Pr[Y_m - Y_0 \ge s] \le \exp\left(-\frac{s^2/2}{\sigma^2 + Rs/3}\right). \tag{3}$$

Plugging in $R = 1$, $\sigma^2 = pm$, $s = O(k \ln N \sqrt{\alpha t n})$ gives

$$\frac{s^2/2}{\sigma^2 + Rs/3} \ge \frac{s^2/2}{pm + s/3} \ge \frac{s^2/2}{\alpha t n + s/3} \ge \Omega(\min(s^2/(\alpha t n), s)) \ge \min(2k^2 \ln^2 N, 2k \ln N \sqrt{\alpha t n}) \ge 2k \ln N.$$

where the second inequality used the fact that $p = 1 - \exp(-t/\overline{\lambda}_k) \le t/\overline{\lambda}_k$ and $m \le \alpha \overline{\lambda}_k n$. Plugging in this bound to (3) proves the lemma. □

How large do we have to set α in Lemma 6? We show that $\alpha := k$ suffices by first applying the graph sparsification routine of Nagamochi and Ibaraki, reducing its number of edges to at most $\lambda_k n$ while maintaining all minimum k-cuts.

Theorem 8 (Nagamochi-Ibaraki [NI92]). *Given an unweighted graph G and parameter λ, there exists a subgraph H with at most λn edges such that all k-cuts of size $\le \lambda$ are preserved. More formally, all sets S with $|\partial_G S| \le \lambda$ satisfy $|\partial_G S| = |\partial_H S|$.*

Proof. For $i = 1, 2, \ldots, \lambda$, let F_i be a maximal forest in $G \setminus \bigcup_{j<i} F_j$. For any edge (u, v) in $G - H = G \setminus \bigcup_i F_i$, there must be an (u, v) path in each F_i, otherwise we would have added edge (u, v) to F_i. These λ paths, along with edge (u, v), imply that every cut that separates u and v has size $\ge \lambda + 1$. Therefore, u and v must lie in the same component of any k-cut of size $\le \lambda$, so removing edge (u, v) cannot affect any such k-cut. □

With Theorem 8 in hand, we now prove the following corollary which we will use in the next section, which combines Lemma 6 and the expectation statement of Lemma 4.

Corollary 9. *Let $t \ge \Omega(1)$ and $N \ge n$ be parameters. Suppose G has at most βn many cuts with weight in the range $[\overline{\lambda}_k, \gamma \overline{\lambda}_k)$ for some constant $1 \le \gamma < 2$. Suppose we contract every edge in G with independent probability $1 - e^{-t/\overline{\lambda}_k}$. Then, with probability at least $1 - N^{-2k}$, number of vertices in the contracted graph is at most $O(\frac{\beta+1}{2-\gamma})ne^{-(\gamma/2)t} + \frac{\gamma}{2}(k-1)t + O(k \ln N \sqrt{ktn})$.*

Proof. First, apply Theorem 8 to the input graph G, obtaining a graph H of at most $\lambda_k n$ edges with the same minimum k-cut value λ_k. We can imagine contracting the graph G by first contracting each edge in H with independent probability $1 - e^{-t/\overline{\lambda}_k}$, and then contracting each edge in $G - H$ with the same probability. Applying Lemma 4 and Lemma 6 with $\alpha := k$ on the graph H, we obtain that contracting the edges in H alone gives us at most $O(\frac{\beta+1}{2-\gamma})ne^{-(\gamma/2)t} + \frac{\gamma}{2}(k-1)t + O(k \ln N \sqrt{ktn})$ with probability at least $1 - N^{-2k}$. Contracting the edges in $G - H$ afterwards can only reduce the number of remaining vertices, so we are done. □

2.3 Recursion

In this section, we finish the proof of the main theorem, restated below:

Theorem 1 (Main). *Given a graph G and a parameter $k \ge 2$, the Karger-Stein algorithm outputs any fixed minimum k-cut in G with probability at least $n^{-k} \cdot (k \ln n)^{-O(k^2 \ln \ln n)}$.*

We will proceed by a recursive analysis: Lemma 4 (expectation) and Lemma 6 (concentration), packaged together in Corollary 9, show that if we let $t = \frac{1}{2} \ln n$ and contract each edge with probability $p = 1 - e^{t/\overline{\lambda}_k}$, the number of remaining vertices becomes at most $\widetilde{O}(\sqrt{n})$ with high probability. Also note that any fixed minimum k-cut C survives (i.e., no edge in C is contracted) with probability exactly $(1-p)^{\overline{\lambda}_k} = n^{-k/2}$.

We then recursively call Corollary 9 on the contracted graph until the number of vertices becomes smaller than some threshold. Formally, let $n_0 := n$ and $G_0 := G$. In the ith iteration, we set $t_i := \frac{1}{2} \ln n_{i-1}$ and contract each edge of G_{i-1} with probability $p = 1 - e^{t_i/\overline{\lambda}_k}$. The above analysis shows that with probability at least $(n_{i-1})^{-k/2}$, no edge in C is contracted and $n_i \le \widetilde{O}(\sqrt{n_{i-1}})$.

If the second guarantee was precisely $n_i \le \sqrt{n_{i-1}}$, iterating at most $T = \lg \lg n$ steps ensures that $n_T \le O(1)$, and the final probability that C survives at the end is roughly at least $n^{-k/2} \cdot n^{-k/4} \cdot n^{-k/8} \cdots \approx n^{-k}$. When the number of vertices becomes small, the naive Karger-Stein analysis can be applied. The proof below formalizes this intuition and accounts the fact that we can only ensure $n_i \le \widetilde{O}(\sqrt{n_{i-1}})$ in each iteration.

Proof. We prove the theorem by recursively applying Corollary 9 to reduce the number of vertices. Given a graph G with

n vertices, let $\gamma := 2 - 1/\ln n$, and apply the following extremal theorem proved in Section 3.

Theorem 10 (Extremal Theorem). *For any $\gamma < 2$, there are at most $(\max(\frac{1}{2-\gamma}, k))^{O(k)}n$ many cuts with weight less than $\gamma\overline{\lambda}_k$.*

Let $\beta = (\max(\frac{1}{2-\gamma}, k))^{O(k)} = (\max(\ln n, k))^{O(k)}$ so that there are at most βn many cuts with weight less than $\gamma\overline{\lambda}_k$. The parameters β and γ will not change throughout the proof. Fix a minimum k-cut C of G so that $|C| = \lambda_k$.

Let $n_0 = \widehat{n}_0 = n$ and $G_0 = G$. For each $i = 1, 2, \dots, T$ where $T = (\lg\lg n_0 - O(1))$, the ith iteration involves setting parameters $t_i := \frac{1}{2}\ln\widehat{n}_{i-1}$, $\widehat{n}_i := M\widehat{n}_{i-1}^{1/2}$ for some $M = O\left(\frac{\beta+1}{2-\gamma}\right)$ to be determined later, and contracting each edge in G_{i-1} with probability $p_i := 1 - e^{-t_i/\overline{\lambda}_k}$ to obtain G_i. Let n_i be the number of vertices of G_i. In each iteration i, we want to ensure that the following events happen with high probability in each iteration, given that the same events happened in the previous iteration.

1. No edge in C is contracted.
2. $n_i \le \widehat{n}_i$.

For the event 1, the probability that no edge in C is contracted is exactly

$$(1-p_i)^{\lambda_k} = e^{-kt_i} = (\widehat{n}_{i-1})^{-k/2}.$$

For the event 2, we use Corollary 9 on G_{i-1} with parameter $N \leftarrow \widehat{n}_{i-1}$. Since C is still a minimum k-cut of G_{i-1}, the minimum k-cut value of G_{i-1} is still $\overline{\lambda}_k$. Applying Corollary 9 to G_i with same β and γ ensures that with probability at least $1 - (\widehat{n}_{i-1})^{-2k}$, n_i is at most

$$O\left(\frac{\beta+1}{2-\gamma}\right) n_{i-1}\exp\left(-\frac{\gamma}{2}t_i\right) + \frac{\gamma}{2}(k-1)t_i + O\left(k\ln\widehat{n}_{i-1}\sqrt{kt\widehat{n}_{i-1}}\right).$$

The last term in the above expression is at most $O\left(\frac{\beta+1}{2-\gamma}\right)(\widehat{n}_{i-1})^{1/2}$ using the fact that $(k\ln\widehat{n}_{i-1}\sqrt{kt}) \le O((\beta+1)/(2-\gamma))$. The first two terms can be upper bounded by

$$\begin{aligned} & O\left(\frac{\beta+1}{2-\gamma}\right) n_{i-1}\exp\left(-\frac{\gamma}{2}t_i\right) + \frac{\gamma}{2}(k-1)t_i \\ \le\ & O\left(\frac{\beta+1}{2-\gamma}\right) n_{i-1}\exp\left(-\frac{\gamma}{2}t_i\right) && (4)\\ \le\ & O\left(\frac{\beta+1}{2-\gamma}\right) n_{i-1}(\widehat{n}_{i-1})^{-\gamma/4} && (5)\\ \le\ & O\left(\frac{\beta+1}{2-\gamma}\right)(\widehat{n}_{i-1})^{1/2} \quad , && (6)\end{aligned}$$

where (4) uses $\frac{\gamma}{2}(k-1)t_i \le k\ln n_0 \le O(\frac{\beta+1}{2-\gamma})$, (5) uses the definition of t_i, and (6) uses $(\widehat{n}_{i-1})^{-\gamma/4} = (\widehat{n}_{i-1})^{-1/2+1/4(\ln n_0)} = O(\widehat{n}_{i-1})$. It follows that with probability at least $1-(\widehat{n}_{i-1})^{-2k}$, we have

$$n_i \le M(\widehat{n}_{i-1})^{1/2} = \widehat{n}_i \tag{7}$$

for large enough $M = O\left(\frac{\beta+1}{2-\gamma}\right)$.

Taking a union bound, the probability that events 1 and 2 both happen is at least

$$(\widehat{n}_{i-1})^{-k/2} - (\widehat{n}_{i-1})^{-2k} = (\widehat{n}_{i-1})^{-k/2}\cdot(1-(\widehat{n}_{i-1})^{-3k/2}).$$

Let $\widehat{x}_i := \lg\widehat{n}_i$ for all $i \ge 0$, so that from (7), we obtain

$$\widehat{x}_i \le \lg M + \frac{1}{2}\widehat{x}_{i-1},$$

which implies that

$$\widehat{x}_i \le \lg M \cdot \left(1 + \frac{1}{2} + \frac{1}{4} + \dots + 2^{-i+1}\right) + 2^{-i}\widehat{x}_0 \le 2\lg M + 2^{-i}\widehat{x}_0, \tag{8}$$

so with $T := \lg\widehat{x}_0 - O(1) = \lg\lg n_0 - O(1)$ steps, $\widehat{x}_T = O(\lg M)$, which translates to $\widehat{n}_T = M^{O(1)}$.

We finally compute the probability that both events happen for each $i = 1, \dots, T = \lg\lg n_0 - O(1)$.

$$\begin{aligned} & \prod_{i=1}^{T}\left((\widehat{n}_{i-1})^{-k/2}\cdot(1-(\widehat{n}_{i-1})^{-3k/2})\right) && (9)\\ =\ & \left(\prod_{i=1}^{T}(\widehat{n}_{i-1})^{-k/2}\right)\cdot\left(\prod_{i=1}^{T}(1-(\widehat{n}_{i-1})^{-3k/2})\right) && (10)\end{aligned}$$

For the second product, the recursive definition $\widehat{n}_i := M\widehat{n}_{i-1}^{1/2}$ also implies $\widehat{n}_i \ge n^{2^{-i}}$, so we can choose $O(1)$ in the definition $T = \lg\lg n_0 - O(1)$ to ensure $\widehat{n}_{T-1} \ge 5$. Then the second product can be shown to be at least $\Omega(1)$, as

$$\left(\prod_{i=1}^{T}(1-(\widehat{n}_{i-1})^{-3k/2})\right) \ge 1 - \sum_{i=1}^{T}(\widehat{n}_{i-1})^{-3k/2},$$

and the sequence $\{(\widehat{n}_i)^{-3k/2}\}$ is at least exponentially increasing with the last term at most 5^{-3} (since $k \ge 2$).

For the first product of (10), we use the fact $\widehat{x}_i \le 2\lg M + 2^{-i}\widehat{x}_0$ to bound

$$\sum_{i=1}^{T}\widehat{x}_{i-1} \le 2T\lg M + 2\widehat{x}_0,$$

which leads to

$$\begin{aligned}\left(\prod_{i=1}^{T}(\widehat{n}_{i-1})^{-k/2}\right) = \left(\prod_{i=1}^{T}(\widehat{n}_{i-1})\right)^{-k/2} &\le 2^{\left(\sum_{i=1}^{T}(\widehat{x}_{i-1})\right)\cdot(-k/2)}\\ &= \left(n^2M^{2T}\right)^{-k/2} = n^{-k}\cdot M^{-Tk}.\end{aligned}$$

Therefore, with probability at least $\Omega(n^{-k}\cdot M^{-Tk})$, events 1 and 2 happen for $i = 1, \dots, T$ which means that no edge in C is contracted and $n_T \le M^{O(1)}$. After this point, we can switch the standard Karger-Stein analysis of the same process where exactly one edge is contracted in each iteration. It shows that if C will be output with probability at least $M^{-O(k)}$. Altogether, the minimum k-cut C survives with probability at least (using $M \le (\max(\ln n, k))^{O(k)}\cdot\ln n \le (k\ln n)^{O(k)}$ and $T \le \lg\lg n$),

$$n^{-k}\cdot M^{-O(Tk)} = n^{-k}\cdot(k\ln n)^{-O(k^2\ln\ln n)}.$$

This completes the proof. □

2.4 Karger-Stein Recursive Contraction

We observe that the recursive procedure of the previous section (Section 2.3) can be "algorithmized" to obtain an algorithm similar to the Karger-Stein recursive contraction algorithm [KS96], which proves Corollary 3. Let the input to the recursive algorithm be (G, i), where G is the input graph and i is the iteration number from the previous section. We start with $(G, 0)$ and always ensure that G has at most $\widehat{n}_{i-1}$ vertices. For an input (G, i), for $(\widehat{n}_{i-1})^{k/2} \cdot (1-(\widehat{n}_{i-1})^{3k/2}$ independent trials, we contract each edge with probability $p_i := 1 - e^{-t_i/\overline{\lambda}_k}$, and if the contracted graph H has at most $\widehat{n}_i$ vertices, we recursively solve $(H, i+1)$. Let us first assume that the algorithm knows λ_k (and hence $\overline{\lambda}_k$ as well).

We have shown in the previous section that the probability that a fixed k-cut in G of size λ_k survives on iteration i is at least

$$(\widehat{n}_{i-1})^{-k/2} \cdot (1 - (\widehat{n}_{i-1})^{-3k/2}).$$

Since we make $(\widehat{n}_{i-1})^{k/2} \cdot (1 - (\widehat{n}_{i-1})^{3k/2})$ independent trials, with probability at least a constant $c > 0$, the k-cut survives in at least one trial. Therefore, the probability that the fixed k-cut survives throughout a full path down the recursion tree is at least c^T, where $T = \lg\lg n_0 - O(1)$ is the number of recursion levels. In other words, the recursive contraction algorithm outputs a minimum k-cut with probability at least $e^{-O(\lg\lg n_0)}$.

We now bound the running time. For iteration i, the algorithm makes

$$\prod_{j=1}^{i} (\widehat{n}_{j-1})^{k/2} \cdot (1 - (\widehat{n}_{j-1})^{3k/2}) \le O(1) \cdot \prod_{j=1}^{i-1} (\widehat{n}_{j-1})^{k/2}$$

many recursive instances (G, i), and contracting the edges in each instance takes $C\widehat{n}_{i-1}^2$ time for some constant $C > 0$, for a total time of

$$T(i) := \prod_{j=1}^{i} (\widehat{n}_{j-1})^{k/2} \cdot C\widehat{n}_{i-1}^2.$$

Let us first assume that $k \ge 4$. With this assumption, we now claim that $T(i)$ is geometrically increasing in i. Recall that we defined $\widehat{n}_i = M\widehat{n}_{i-1}^{1/2}$ for some $M = \omega(1)$. Since $k \ge 4$, we have $\widehat{n}_{i-1}^2 = \left(\frac{1}{M}\widehat{n}_i\right)^4 = \frac{1}{M^4}\widehat{n}_i^2 \cdot \widehat{n}_i^2 \le \frac{1}{M^4}\widehat{n}_i^{k/2} \cdot \widehat{n}_i^2$, we have

$$T(i+1) = \prod_{j=1}^{i+1} (\widehat{n}_{j-1})^{k/2} \cdot C\widehat{n}_i^2 = \prod_{j=1}^{i} (\widehat{n}_{j-1})^{k/2} \cdot \widehat{n}_i^{k/2} \cdot C\widehat{n}_i^2$$

$$\ge M^4 \prod_{j=1}^{i} (\widehat{n}_{j-1})^{k/2} \cdot C\widehat{n}_{i-1}^2 = M^4 \cdot T(i).$$

Therefore, the total running time $\sum_{i=1}^{T} T(i)$ is dominated by its last term

$$\prod_{i=1}^{T} (\widehat{n}_{i-1})^{-k/2} \cdot M^{O(1)} \overset{(??)}{\le} n^k \cdot M^{Tk} = 2^{O(\ln\ln n)^2} n^k,$$

as desired.

We now handle the cases $k \le 3$. Here, the first term $T(1)$ in the recursion becomes $\widehat{n}_0^{k/2} \cdot C\widehat{n}_0^2 = Cn^{2+k/2}$, which is already larger than n^k by a polynomial factor. Hence, we need a different recursion procedure. Karger and Stein's original recursive analysis [KS96] works for $k = 2$. For $k = 3$, instead of defining $\widehat{n}_i \approx \widehat{n}_{i-1}^{1/2}$, we set $\widehat{n}_i \approx \widehat{n}_{i-1}^{2/3}$ instead, so that the new running times $T(i)$ become geometrically increasing as before. Since carrying out the new analysis will require us to redo Section 2.2, we omit it for brevity.

Finally, address the issue of the algorithm not knowing $\overline{\lambda}_k$. The algorithm first computes a $(2+\epsilon)$-approximation in near-linear time [Qua18] and guesses a value $\tilde{\lambda}_k \in [\overline{\lambda}_k, 2\overline{\lambda}_k]$. It runs the contraction procedure with polylog(n) guesses for $\overline{\lambda}_k$ such that one of them, say $\overline{\lambda}'_k$, is within the range $[(1 - \frac{1}{\text{polylog}(n)})\overline{\lambda}_k, \overline{\lambda}_k]$. It is not hard to see that if the algorithm sets $p_i := 1 - e^{-t_i/\overline{\lambda}'_k}$ instead, the running time and success probability do not change asymptotically. For example, the probability that a fixed k-cut survives is now $\exp(-t_i\lambda_k/\overline{\lambda}'_k) \ge \exp(-t_i k(1 - \frac{1}{\text{polylog}(n)}))$, and the $(1 - \frac{1}{\text{polylog}(n)})$ factor can be shown to be negligible.

3 CUTS AND SUNFLOWERS

In this section, we prove that for any $\gamma < 2$, every graph has a small number of cuts whose weight is $\gamma\overline{\lambda}_k$. Our main result in this section is:

Theorem 10 (Extremal Theorem). *For any $\gamma < 2$, there are at most $(\max(\frac{1}{2-\gamma}, k))^{O(k)} n$ many cuts with weight less than $\gamma\overline{\lambda}_k$.*

3.1 The Sunflower Lemma, and Refinements

Recall that given a set system $\mathcal{F}$ over a universe U, an *r-sunflower* is a collection of r subsets $F_1, \ldots, F_r \in \mathcal{F}$ such that their pairwise intersection is the same: there exists a *core* $S \subseteq U$ such that $F_i \cap F_j = S$ for all i, j, and hence $\cap_i F_i = S$. Let sf(d, r) be the smallest number such that any set system with n elements and more than sf(d, r) sets of cardinality d must have an r-sunflower. The classical bound of Erdős and Rado [ER60] shows that sf$(d, r) \le d!(r-1)^d$. A recent breakthrough by Alweiss et al. [ALWZ19] proves that sf$(d, r) \le (\lg d)^d (r \cdot \lg\lg d)^{O(d)}$.

Corollary 11. *Let $\mathcal{F}$ be a family of sets over some universe, where every set has size at most d. If $|\mathcal{F}| > (d+1) \cdot$ sf(d, r), then $\mathcal{F}$ contains an r-sunflower.*

Proof. Group the sets in $\mathcal{F}$ by their sizes, which range from 0 to d. For some $d' \in [0, d]$, there are more than

$$\frac{1}{d+1} \cdot \big((d+1)\,\mathrm{sf}(d, r)\big) \ge \mathrm{sf}(d', r)$$

sets of size exactly d', since sf(d, r) is monotone in d. The result follows from applying the definition of sf(d', r) on the sets in $\mathcal{F}$ of size d'. □

For our applications for cuts, we want a sunflower with nonempty core. In this case, the bound must depend on the size of the universe n, since the set system with n singleton sets does not contain a sunflower with nonempty core. The following lemma proves that the above bound, multiplied by $\approx nd$, can guarantee a sunflower with nonempty core.

Lemma 12. *Let $\mathcal{F}$ be a family of sets over a universe of n elements, where every set has size at most d. If $|\mathcal{F}| > (d+2)\,$sf$(d, r)n$, then $\mathcal{F}$ contains an r-sunflower with nonempty core.*

Proof. We prove the contrapositive: suppose that $\mathcal{F}$ does not have an r-sunflower with nonempty core. For each element $v \in U$, consider the set $\mathcal{F}_v := \{F \in \mathcal{F} : F \ni v\}$. If there exists an r-sunflower in $\mathcal{F}_v$ for any $v \in U$, then this sunflower has a nonempty core (since the core contains v), contradicting our assumption. Therefore, by Corollary 11, $|\mathcal{F}_v| \le (d+1) \cdot \mathrm{sf}(d,r)$ for each $v \in U$. Every set in $\mathcal{F}$ is included in some $\mathcal{F}_v$ except possibly $\varnothing$, so

$$|\mathcal{F}| \le (d+1)\,\mathrm{sf}(d,r) \cdot n + 1 \le (d+2)\,\mathrm{sf}(d,r)n,$$

proving the contrapositive. □

Additionally we want multiple sunflowers, each with distinct, nonempty nonempty core. Note that the sunflower cores may intersect, even though they are distinct. The following lemma shows we can also achieve this.

Lemma 13. *Let $\mathcal{F}$ be a family of sets over a universe of n elements, where every set has size at most d. If $|\mathcal{F}| > s(d+2)\mathrm{sf}(d,r)n$, then $\mathcal{F}$ contains s many r-sunflowers, each with distinct, nonempty cores.*

Proof. We iteratively construct s sunflowers with distinct cores. Initialize $\mathcal{F}' := \mathcal{F}$, and on each iteration, consider a maximal set C such that there exists an r-sunflower in $\mathcal{F}'$ with core C. Inductively we ensure that such a set $C \neq \varnothing$ exists; this holds for the base case by Lemma 12.

Moreover, we claim that the set $\mathcal{F}'_C := \{F \in \mathcal{F}' : F \supseteq C\}$ has size at most $(d+2)\mathrm{sf}(d,r)n$. Indeed, if not, then applying Lemma 12 on the set system $\{F \setminus C : F \in \mathcal{F}', F \supseteq C\}$ (which has the same cardinality as $\mathcal{F}'$), we obtain an r-sunflower with sets $S_1, \dots, S_r$ and nonempty core C'. Then, the sets $S_1 \cup C, \dots, S_r \cup C \in \mathcal{F}'$ form an r-sunflower with core $C \cup C'$, contradicting the maximality of the set C.

We now remove the sets in $\mathcal{F}'_C$ from $\mathcal{F}'$ (i.e., update $\mathcal{F}' \leftarrow \mathcal{F}' \setminus \mathcal{F}'_C$). Now the core on any subsequent iteration cannot be C, since we have removed all the sets that contained C. The size of $\mathcal{F}'$ drops by at most $|\mathcal{F}'_C| \le (d+2)\,\mathrm{sf}(d,r)n$ each iteration, so if $|\mathcal{F}| > s(d+2)\,\mathrm{sf}(d,r)n$ to begin with, then we can proceed for s iterations, obtaining s many r-sunflowers with distinct, nonempty cores. □

3.2 Removing the Size Restriction: Venn Diagrams

The above sunflower lemmas proved that a sunflower-free set system $\mathcal{F}$ must have few sets, as long as each set in the system has bounded size. The following lemma replaces the assumption on the bounded size by the assumptions that (a) every k sets in the system have small number of occupied regions in their Venn diagram, and (b) the set system of the complements of the sets do not contain many sunflowers either.

To make this formal, we introduce some notation. Given k sets $F_1, \dots, F_k$, we denote their *Venn diagram* by $\mathrm{Venn}(F_1, \dots, F_k)$. An *atom* denotes a nonempty region of the diagram. Formally, an atom is a nonempty set that can be expressed as $G_1 \cap \dots \cap G_k$, where for each i, the set G_i is either the set F_i, or its complement $\overline{F_i}$. Also, let $\overline{\mathcal{F}} := \{\overline{F} : F \in \mathcal{F}\}$ be the collection of complements of the sets in $\mathcal{F}$.

Lemma 14. *Let $\mathcal{F}$ be a set system on n elements satisfying the following:*

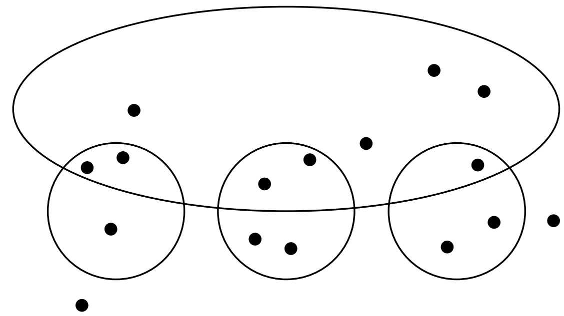

Figure 2: The Venn diagram above has eight atoms.

i. *For every k sets $F_1, \dots, F_k \in \mathcal{F}$, their Venn diagram $\mathrm{Venn}(F_1, \dots, F_k)$ has less than $2k$ atoms.*
ii. *Each of $\mathcal{F}$ and $\overline{\mathcal{F}}$ does not contain s many r-sunflowers, each with distinct, nonempty cores.*

Then, $|\mathcal{F}| \le 10s \cdot k(5k+2) \cdot \mathrm{sf}(5k,r) \cdot n$.

Proof. For fixed r, k, s, let $\mathrm{ex}(n)$ (ex for extremal) be the maximum size of a set $\mathcal{F}$ on n elements satisfying conditions (i) and (ii). We prove by induction on n that

$$\mathrm{ex}(n) \le 10s \cdot k(5k+2) \cdot \mathrm{sf}(5k,r) \cdot \max\{1, n-4k\},$$

with the base cases $n \le 5k$.

Base case: $n \le 5k$. In this case, each set has size at most $n \le 5k$, so using Lemma 13, so the number of sets in $\mathcal{F}$ is at most

$$s(5k+2)\mathrm{sf}(5k,r)\cdot 5k \le 10s\cdot k(5k+2)\cdot\mathrm{sf}(5k,r)\cdot\max\{1,n-4k\} = \mathrm{ex}(n).$$

Inductive step: $n > 5k$. First, suppose that every set $F \in \mathcal{F}$ satisfies either $|F| \le 5k$ or $|F| \ge n-5k$. By Lemma 13 on $\mathcal{F}$ and $\overline{\mathcal{F}}$ respectively, there are at most $s(5k+2) \cdot \mathrm{sf}(5k,r) \cdot n$ many sets of size at most $5k$, and also at most $s(5k+2) \cdot \mathrm{sf}(5k,r) \cdot n$ many sets of size at least $n-5k$. Applying the bound $n \le 5(n-4k)$ and using that $k \ge 1$, we obtain

$$\begin{aligned} |\mathcal{F}| &\le 2s(5k+2) \cdot \mathrm{sf}(5k,r) \cdot n \\ &\le 2s \cdot k(5k+2) \cdot \mathrm{sf}(5k,r) \cdot 5\max\{1, n-4k\} = \mathrm{ex}(n), \end{aligned}$$

as desired.

Otherwise, there exists a set S with $5k < |S| < n-5k$. For $i = 1, 2, \dots, k-1$, while there exists a set $T_i \in \mathcal{F}$ such that the Venn diagram $\mathrm{Venn}(S, T_1, T_2, \dots, T_i)$ on the $i+1$ sets contains at least $2(i+1)$ atoms, choose an arbitrary such set T_i. Suppose this process continues until the index i reaches value $\ell \in [k-1]$. If $\ell = k-1$, then $\mathrm{Venn}(S, T_1, \dots, T_\ell)$ is composed of k sets and has at least $2k$ atoms, which cannot happen by assumption. Therefore, $\ell < k-1$. We say that a set F *cuts* another set F' if both the regions $F \cap F'$ and $F' \setminus F$ are non-empty. By our stopping condition, every set $F \in \mathcal{F}$ cuts at most one atom in $\mathrm{Venn}(S, T_1, \dots, T_\ell)$; indeed, if a set F cuts two atoms or more, we would have added it as $T_{\ell+1}$ and continued.

Let the atoms of $\mathrm{Venn}(S, T_1, \dots, T_\ell)$ inside S be $A_1, \dots, A_i$, so that $A_1 \cup \dots \cup A_i = S$; and let the atoms outside S be $B_1, \dots, B_j$, so that $B_1 \cup \dots \cup B_j = U \setminus S$. Define two new collections of elements

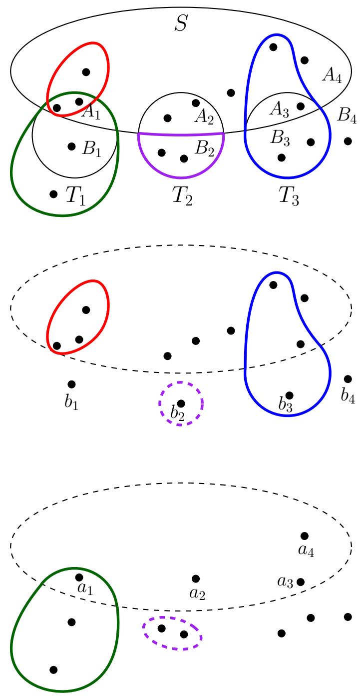

Figure 3: Construction of the set systems $(X, \mathcal{F}_a)$ (middle) and $(Y, \mathcal{F}_b)$ (bottom) given the set system on the top and sets $S, T_1, T_2, T_3 \in \mathcal{F}$. Each colored set represents another set in $\mathcal{F}$ which is added to either $(X, \mathcal{F}_a)$ or $(Y, \mathcal{F}_b)$. The red and blue sets must be added to $(X, \mathcal{F}_a)$, and the green set must be added to $(Y, \mathcal{F}_b)$. The purple set can be added to either $(X, \mathcal{F}_a)$ or $(Y, \mathcal{F}_b)$.

$E_a := \{a_1, \dots, a_i\}$ and $E_b := \{b_1, \dots, b_j\}$, and define $X := S \cup E_b$ and $Y := (U \setminus S) \cup E_a$. We build two set systems $(X, \mathcal{F}_a)$ and $(Y, \mathcal{F}_b)$ as follows (see Figure 3). Initialize $\mathcal{F}_a = \mathcal{F}_b := \varnothing$; for each set $F \in \mathcal{F}$, we have three cases:

(1) If F cuts an atom A_h inside S, then add the set $(F \cap S) \cup \{b_{j'} \mid j' \in [j], F \supseteq B_h\}$ into $\mathcal{F}_a$.
(2) Else, if F cuts an atom B_h outside S, then add the set $(F \setminus S) \cup \{a_{i'} \mid i' \in [i], F \supseteq A_h\}$ into $\mathcal{F}_b$.
(3) Else, F does not cut any atom. Execute either step (1) or step (2).

Here's another equivalent way to look at this process. For $\mathcal{F}_a$, we can think taking the set system $(U, \mathcal{F})$, removing the sets that cut an atom outside S, and then contracting the atoms $B_1, \dots, B_j$ into $b_1, \dots, b_j$, respectively. We can also think of $\mathcal{F}_b$ analogously, by throwing away the sets that cut atoms inside S, and then contacting atoms $A_1, \dots, A_i$. Through this contraction viewpoint, it is clear that if the set system $(U, \mathcal{F})$ satisfy conditions (i) and (ii), then so do the set systems $(X, \mathcal{F}_a)$ and $(Y, \mathcal{F}_b)$. Moreover, since $5k < |S| < n - 5k$, we have

$$|X| = |S| + |E_b| \le |S| + 2k \le (n - 5k) + 2k < |U|$$

and

$$|Y| = |V \setminus S| + |E_a| \le n - |S| + 2k \le (n - 5k) + 2k < |U|,$$

so we can apply induction on n, obtaining

$$\begin{aligned}
|\mathcal{F}| &= |\mathcal{F}_a| + |\mathcal{F}_b| \\
&\le \mathrm{ex}(n - |S| + 2k) + \mathrm{ex}(|S| + 2k) \\
&\le 10s \cdot k(5k+2) \cdot \mathsf{sf}(5k, r) \cdot \\
&\qquad \Big(\max\{1, (n - |S| + 2k) - 4k\} + \max\{1, (|S| + 2k) - 4k\} \Big) \\
&= 10s \cdot k(5k+2) \cdot \mathsf{sf}(5k, r) \cdot \\
&\qquad \Big((n - |S| + 2k) - 4k + (|S| + 2k) - 4k \Big) \\
&= 10s \cdot k(5k+2) \cdot \mathsf{sf}(5k, r) \cdot (n - 4k) = \mathrm{ex}(n),
\end{aligned}$$

completing the induction. □

Later when we apply the above lemma to k-cut, the number of atoms becomes k, so it is sufficient for even k. For odd k, we can slightly strengthen Lemma 14 as follows.

Corollary 15. *Let $\mathcal{F}$ be a set system on n elements satisfying the following:*

i. *There do not exist sets $S_1, \dots, S_{k-1}$ such that $\mathrm{Venn}(S_1, \dots, S_{k-1})$ has at least $2(k-1) + 1$ atoms.*
i'. *There do not exist sets $S_1, \dots, S_k$ such that $\mathrm{Venn}(S_1, \dots, S_{k-1})$ has exactly $2(k-1)$ atoms, and the set S_k cuts at least two atoms in $\mathrm{Venn}(S_1, \dots, S_{k-1})$.*
ii. *Each of $\mathcal{F}$ and $\overline{\mathcal{F}}$ does not contain s many r-sunflowers, each with distinct, nonempty cores.*

Then, $|\mathcal{F}| \le 10s \cdot k(5k+2) \cdot \mathsf{sf}(5k, r) \cdot n$.

Proof. The proof is identical; the only additional observation is that when we iteratively construct $S, T_1, \dots, T_\ell$ for $\ell \le k - 1$, observe that the set T_ℓ cuts at least two atoms of $\mathrm{Venn}(S, T_1, \dots, T_{\ell-1})$ by construction. In particular, if the construction continued until $\ell = k - 1$, then either the sets $S, T_1, \dots, T_{k-2}$ violate condition (i), or the sets $S, T_1, \dots, T_{k-1}$ violate condition (i'). Therefore, every time we carry out this process, we must stop at $\ell < k - 1$. When we stop, the condition (i), though it is slightly more relaxed than the condition (i) of Lemma 14, still ensures that $|E_a|, |E_b| \le 2k$, so the same inductive argument works. □

3.3 Relating Cuts and Sunflowers

Recall that $\overline{\lambda}_k$ is the size of the minimum k-cut divided by k, and $\gamma \in [1, 2)$ is a fixed parameter. In this section, we use the previous tools for sunflowers to bound the number of small cuts (of size $\le \gamma \overline{\lambda}_k$) in a graph. First, the following lemma, independent of sunflowers, shows that there cannot be many tiny cuts (of size $< \overline{\lambda}_k$) in a graph.

Lemma 16. *There are at most 2^{k-1} many cuts with weight less than $\overline{\lambda}_k$.*

Proof. Suppose, otherwise, that there are more than 2^{k-1} sets; let $\mathcal{S}$ be the collection of these sets. We will iteratively construct a k-cut of size less than $k\overline{\lambda}_k = \lambda_k$ contradicting the definition of λ_k, the size of the minimum k-cut.

Begin with an arbitrary set $S_1 \in \mathcal{S}$, and while $\mathrm{Venn}(S_1, \ldots, S_{i-1})$ has less than k components, choose an arbitrary set $S_i \in \mathcal{S}$ such that $\mathrm{Venn}(S_1, \ldots, S_i)$ has at least one more component than $\mathrm{Venn}(S_1, \ldots, S_{i-1})$. We show that such a set S_i always exists. Let $A_1, \ldots, A_\ell$ be the atoms of $\mathrm{Venn}(S_1, \ldots, S_{i-1})$; the only sets $T \in \mathcal{S}$ such that $\mathrm{Venn}(S_1, \ldots, S_{i-1}, T)$ has the same number of components as $\mathrm{Venn}(S_1, \ldots, S_{i-1})$ are sets of the form $\bigcup_{i \in I} A_i$ for some subset $I \subseteq [\ell]$. Since there are at most $2^\ell \le 2^{k-1}$ such sets and $|\mathcal{S}| > 2^{k-1}$, a satisfying set S_i always exists.

At the end, we have at most $k-1$ sets $S_1, \ldots, S_i$ such that $\mathrm{Venn}(S_1, \ldots, S_i)$ has at least k components. Therefore, the edge set $\partial S_1 \cup \cdots \cup \partial S_i$ is a k-cut, and it has weight less than $i\overline{\lambda}_k < k\overline{\lambda}_k = \lambda_k$, achieving the desired contradiction. □

Finally, the following lemma proves that many sunflowers consisting of cuts of size $\le \gamma\overline{\lambda}_k$ will lead to a better k-cut than $k\overline{\lambda}_k$, leading to contradiction.

Lemma 17. *Fix a constant $1 \le \gamma < 2$, and let $\mathcal{F}$ be the family of sets $\{S \subseteq V : w(\partial S) \le \gamma\overline{\lambda}_k\}$. Then, for any $r > \frac{2\gamma}{2-\gamma} + 1$, both $\mathcal{F}$ and $\overline{\mathcal{F}}$ do not contain 2^k many $(r+k-2)$-sunflowers with distinct, nonempty cores.*

Proof. Since $w(\partial S) \le \gamma\overline{\lambda}_k \iff w(\partial(V \setminus S)) \le \gamma\overline{\lambda}_k$, we have $\mathcal{F} = \overline{\mathcal{F}}$, so it suffices to only consider $\mathcal{F}$. Suppose, otherwise, that there are 2^k many $(r+k-2)$-sunflowers with distinct, nonempty cores. Let $\mathcal{F}_{\mathrm{small}} := \{S \mid \varnothing \subsetneq S \subsetneq V,\ w(\partial S) < \overline{\lambda}_k\}$, so that Lemma 16 implies that $|\mathcal{F}_{\mathrm{small}}| \le 2^{k-1}$. Then, there must exist at least one sunflower in this collection whose core does not belong to $\mathcal{F}_{\mathrm{small}}$. Let $S_1, \ldots, S_{r+k-2} \in \mathcal{F}$ be the sets of this $(r+k-2)$-sunflower with petals $P_i := S_i \setminus \bigcup_{j \ne r} S_j$ and nonempty core $C := \bigcap_i S_i \notin \mathcal{F}_{\mathrm{small}}$. Since the petals P_i are disjoint, at most $k-2$ of them are in $\mathcal{F}_{\mathrm{small}}$, since otherwise, we get $k-1$ disjoint sets in $\mathcal{F}_{\mathrm{small}}$ which together form a k-cut with weight less than $(k-1)\overline{\lambda}_k < \lambda_k$. Therefore, without loss of generality (by reordering the sets S_i), assume that $P_1, \ldots, P_r \notin \mathcal{F}_{\mathrm{small}}$. Since C and $P_1, \ldots, P_r$ are all cuts in the graph (in particular, $\varnothing \ne C \ne V$ and $\varnothing \ne P_i \ne V$) and are not in $\mathcal{F}_{\mathrm{small}}$, we have $w(\partial C) \ge \overline{\lambda}_k$ and $w(\partial P_i) \ge \overline{\lambda}_k$ for each $i \in [r]$. For each $i \in [r]$, we have

$$\begin{aligned}\gamma\overline{\lambda}_k &\ge w(\partial S_i) = w(\partial(C \cup P_i)) \\ &= w(\partial C) + w(\partial P_i) - 2w(E[C, P_i]) \ge 2\overline{\lambda}_k - 2w(E[C, P_i]),\end{aligned}$$

so $w(E[C, P_i]) \ge (2-\gamma)\overline{\lambda}_k/2$. Now observe that the edges in $E[C, P_i]$ for $i = 2, \ldots, r$ are included in $\partial(C \cup P_1)$. It follows that

$$\begin{aligned}\gamma\overline{\lambda}_k &\ge w(\partial(C \cup P_1)) \ge w(E[C, P_2] \cup \cdots \cup E[C, P_r]) \\ &= \sum_{i=2}^{r} w(E[C, P_i]) \ge (r-1)\frac{(2-\gamma)\overline{\lambda}_k}{2},\end{aligned}$$

so $r - 1 \le \frac{2\gamma}{2-\gamma}$, contradicting the assumption that $r > \frac{2\gamma}{2-\gamma} + 1$. □

We are finally ready to prove Theorem 10, which we restate here for convenience.

Theorem 10 (Extremal Theorem). *For any $\gamma < 2$, there are at most $(\max(\frac{1}{2-\gamma}, k))^{O(k)} n$ many cuts with weight less than $\gamma\overline{\lambda}_k$.*

Proof. Let $\mathcal{F}$ be the set of such cuts, and let $k' := \lceil k/2 \rceil$. We first show that if k is even, then condition (i) of Lemma 14 is satisfied when the parameter k in the lemma is k' instead, and if k is odd, then conditions (i) and (i') of Corollary 15 are satisfied, again with k' for the parameter k. Then, we show that condition (ii) of both Lemma 14 and Corollary 15 are satisfied for parameters s and r that we choose later.

First, consider the case when k is even. Suppose, otherwise, that condition (i) of Lemma 14 is false: there are sets $S_1, \ldots, S_{k'}$ such that $\mathrm{Venn}(S_1, \ldots, S_{k'})$ has at least $2k' = k$ atoms. Then, $\partial S_1 \cup \cdots \cup \partial S_{k'}$ is a k-cut with weight $k' \cdot \gamma\overline{\lambda}_k < \frac{k}{2} \cdot 2\overline{\lambda}_k = \lambda_k$, contradicting the definition of λ_k, the minimum k-cut.

Now consider the case when k is odd. If condition (i) of Corollary 15 is false, then there are sets $S_1, \ldots, S_{k'-1}$ such that $\mathrm{Venn}(S_1, \ldots, S_{k'}$ has at least $2(k'-1)-1 = k$ atoms. Then, $\partial S_1 \cup \cdots \cup \partial S_{k'-1}$ is a k-cut with weight $(k'-1) \cdot \gamma\overline{\lambda}_k < \frac{k}{2} \cdot 2\overline{\lambda}_k = \lambda_k$, contradicting the definition of λ_k, the minimum k-cut. Otherwise, if condition (i') of Corollary 15 is false, then the set of edges $\partial S_1 \cup \cdots \cup \partial S_{k'-1}$ is a $(k-1)$-cut with weight $(k'-1) \cdot \gamma\overline{\lambda}_k < \frac{k-1}{2} \cdot 2\overline{\lambda}_k = (k-1)\overline{\lambda}_k$. Let $A_1, \ldots, A_\ell \subseteq V$ be the atoms in $\mathrm{Venn}(S_1, \ldots, S_{k'-1})$ that are cut by $S_{k'}$, with $\ell \ge 2$ by assumption. Since $\partial S_{k'} \cap E[A_i]$ are disjoint for $i \in [\ell]$, there exists one atom A_i such that

$$w(\partial S_{k'} \cap E[A_i]) \le \frac{1}{\ell} w(\partial S_{k'}) \le \frac{1}{\ell} \cdot \gamma\overline{\lambda}_k < \frac{1}{2} \cdot 2\overline{\lambda}_k = \overline{\lambda}_k.$$

Thus, $\partial S_1 \cup \cdots \cup \partial S_{k'-1} \cup (\partial S_{k'} \cap E[A_i])$ is a k-cut with weight less than $(k-1)\overline{\lambda}_k + \overline{\lambda}_k = \overline{\lambda}_k$, a contradiction. Thus, conditions (i) and (i') of Corollary 15 are satisfied.

Now, fix the parameters $s := 2^k$, $r := \lceil \frac{2\gamma}{2-\gamma} + 2 \rceil$, $r' := r + k - 2$, and $k' := \lceil k/2 \rceil$. By Lemma 17, both $\mathcal{F}$ and $\overline{\mathcal{F}}$ do not contain s many r'-sunflowers with distinct, nonempty cores, fulfilling condition (ii) of Lemma 14 and Corollary 15 (with r' in place of r). Therefore, by Lemma 14 or Corollary 15 when k is even or odd respectively, and using $\mathrm{sf}(d, r) \le (\lg d)^d (r \cdot \lg\lg d)^{O(d)} \le (r \lg d \lg\lg d)^{O(d)}$ [ALWZ19],

$$\begin{aligned}|\mathcal{F}| &\le 10s \cdot k'(5k'+2) \cdot \mathrm{sf}(5k', r') \cdot n \\ &\le 2^{O(k)} \cdot (r' \lg k' \lg\lg k')^{O(k')} n \\ &= 2^{O(k)} \cdot \left(\left(\lceil \frac{2\gamma}{2-\gamma} \rceil + k\right) \lg k' \lg\lg k'\right)^{O(k')} \cdot n \\ &\le \left(\max\left(\frac{1}{2-\gamma}, k\right)\right)^{O(k)} \cdot n\end{aligned}$$

as desired. As an aside, using the classical Erdős-Rado bound [ER60] of $\mathrm{sf}(d, r) \le d!(r-1)^d$ gives the same result, up to a constant in the $O(k)$ exponent, as does the conjectured optimal bound of $\mathrm{sf}(d, f) \le (Cr)^d$. □

A OMITTED PROOFS

Proof of Claim 5. We have

$$\tilde{f}(0) = \frac{1}{1-A} n(1-A) = n = f(0)$$

and

$$\tilde{f}'(t) = \frac{1}{1-A} n(-\frac{\gamma}{2} e^{-(\gamma/2)t} + Ae^{-t})$$

$$
\begin{aligned}
&= \frac{1}{1-A} n \left(-\frac{\gamma}{2} e^{-(\gamma/2)t} + \frac{\gamma}{2} A e^{-t} + \left(A - \frac{\gamma}{2} A \right) e^{-t} \right) \\
&= -\frac{\gamma}{2} \cdot \frac{1}{1-A} n (e^{-(\gamma/2)t} - A e^{-t}) + \frac{A - (\gamma/2)A}{1-A} e^{-t} n \\
&= -\frac{\gamma}{2} \tilde{f}(t) + \frac{A - (\gamma/2)A}{1-A} e^{-t} n. \\
&= -\frac{\gamma}{2} \tilde{f}(t) + B e^{-t} n,
\end{aligned}
$$

where the last equality holds because

$$
\begin{aligned}
A = \frac{B}{B + 1 - \gamma/2} &\iff BA + A - \frac{\gamma}{2} A = B \\
&\iff A - \frac{\gamma}{2} A = B - BA = B(1-A) \\
&\iff \frac{A - (\gamma/2)A}{1-A} = B.
\end{aligned}
$$

Since $\tilde{f}(0) = f(0)$ and $\tilde{f}(t)$ satisfies (2) with equality, we have $\tilde{f}(t) \geq f(t)$ for all $t \geq 0$. □

REFERENCES

[ALWZ19] Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang. Improved bounds for the sunflower lemma. *arXiv preprint arXiv:1908.08483*, 2019.

[CQX18] Chandra Chekuri, Kent Quanrud, and Chao Xu. LP relaxation and tree packing for minimum k-cuts. In *2nd Symposium on Simplicity in Algorithms (SOSA 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[ER60] P. Erdős and R. Rado. Intersection theorems for systems of sets. *J. London Math. Soc.*, 35:85–90, 1960.

[Fre75] David A Freedman. On tail probabilities for martingales. *the Annals of Probability*, 3(1):100–118, 1975.

[GH94] Olivier Goldschmidt and Dorit S. Hochbaum. A polynomial algorithm for the k-cut problem for fixed k. *Math. Oper. Res.*, 19(1):24–37, 1994.

[GLL18] Anupam Gupta, Euiwoong Lee, and Jason Li. Faster exact and approximate algorithms for k-cut. In *Foundations of Computer Science (FOCS), 2018 IEEE 59th Annual Symposium on*, 2018.

[GLL19] Anupam Gupta, Euiwoong Lee, and Jason Li. The number of minimum k-cuts: improving the karger-stein bound. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 229–240. ACM, 2019.

[KS96] David R. Karger and Clifford Stein. A new approach to the minimum cut problem. *Journal of the ACM (JACM)*, 43(4):601–640, 1996.

[KYN07] Yoko Kamidoi, Noriyoshi Yoshida, and Hiroshi Nagamochi. A deterministic algorithm for finding all minimum k-way cuts. *SIAM J. Comput.*, 36(5):1329–1341, 2006/07.

[Li19] Jason Li. Faster minimum k-cut of a simple graph. In *Foundations of Computer Science (FOCS), 2019 IEEE 60th Annual Symposium on*, 2019.

[NI92] Hiroshi Nagamochi and Toshihide Ibaraki. Computing edge-connectivity in multigraphs and capacitated graphs. *SIAM J. Discrete Math.*, 5(1):54–66, 1992.

[Qua18] Kent Quanrud. Fast and deterministic approximations for k-cut. *arXiv preprint arXiv:1807.07143*, 2018.

[Tho08] Mikkel Thorup. Minimum k-way cuts via deterministic greedy tree packing. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 159–166. ACM, 2008.