1 Cut Sparsifier

Recall that a *cut* $(S, V \setminus S)$ of a graph $G = (V, E_G)$ is a non-trivial partition of the node set V into two sets S and $V \setminus S$ (where non-trivial means that $S \neq \emptyset$ and $S \neq V$). We denote by $E_G(S, V \setminus S)$ the subset of edges crossing this cut (i.e., $E_G(S, V \setminus S) = \{(u, v) \in E \mid u \in S, v \in V \setminus S\}$). In an unweighted graph G we call $|E_G(S, V \setminus S)|$ the size of the cut $(S, V \setminus S)$, and in a weighted graph G we call $w_G(E_G(S, V \setminus S)) = \sum_{e \in E_G(S, V \setminus S)} w_G(e)$ the weight of the cut $(S, V \setminus S)$ The minimum cut size λ_G is the minimum over the sizes of all cuts of G.

Definition 1.1. *A $(1 \pm \epsilon)$-cut sparsifier $H = (V, E_H)$ of an unweighted graph $G = (V, E_G)$ is a reweighted subgraph (with $E_H \subseteq E_G$) such that for every cut $(S, V \setminus S)$ the following holds:*

$$(1 - \epsilon)|E_G(S, V \setminus S)| \leq w_H(E_H(S, V \setminus S)) \leq (1 + \epsilon)|E_G(S, V \setminus S)|$$

where $|E_G(S, V \setminus S)|$ denotes the number of edges crossing the cut $(S, V \setminus S)$ in G and $w_H(E_H(S, V \setminus S))$ denotes the sum of the weights of the edges crossing the cut $(S, V \setminus S)$.

1.1 Basic Sampling Approach

A straightforward idea for sparsifying a graph is to obtain H by sampling each edge with some fixed probability p. By giving weight $\frac{1}{p}$ to the edges in H, each cut is preserved in expectation.

```
E_H ← ∅
foreach e ∈ E do
    With probability p: E_H ← E_H ∪ {e} and w_H(e) ← 1/p
return H = (V, E_H)
```

Lemma 1.2. *If $H = (V, E_H)$ is obtained from an unweighted graph $G = (V, E)$ by the uniform sampling procedure described above, then for each cut $(S, V \setminus S)$, $\mathrm{Ex}[w_H(E_H(S, V \setminus S))] = |E_G(S, V \setminus S)|$.*

Proof. For every edge $e \in E$, let X_e be the binary random variable that has value $w_H(e) = \frac{1}{p_e}$ if e sampled and 0 otherwise. For such a binary random variable, $\mathrm{Ex}[X_e] = w_H(e) \cdot \Pr[X_e = 1] = 1$. Now consider any cut $(S, V \setminus S)$. By linearity of expectation we get

$$\mathrm{Ex}[w_H(E_H(S, V \setminus S))] = \mathrm{Ex}\left[\sum_{e \in E_G(S, V \setminus S)} X_e\right] = \sum_{e \in E_G(S, V \setminus S)} \mathrm{Ex}[X_e] \sum_{e \in E_G(S, V \setminus S)} 1 = |E_G(S, V \setminus S)|$$

□

Bounding this expectation for each cut is however not enough for our purposes. We aim at a probabilistic guarantee that holds for all cuts simultaneously. From the expectation above, we can, via the Markov Bound, only derive a probabilistic guarantee of the form

$$\Pr[w_H(E_H(S, V \setminus S)) > (1 + \epsilon)|E_G(S, V \setminus S)|] < \frac{1}{1 + \epsilon}$$

for each cut $(S, V \setminus S)$ that is (a) only one-sided an (b) too weak to give a guarantee for all (exponentially many) cuts simultaneously, say via a union bound.

A straightforward application of the two-sided Chernoff bound for sums of binary random variables however would give us a suitable guarantee for all *large enough* cuts. We first restate a suitable version of the Chernoff bound for sums of binary random variables.

Theorem 1.3 (Chernoff Bound for Generalized Poisson Trials). *Let $X_1, \ldots, X_n \in \{0, R\}$ be independent binary random variables (i.e., $X_i = 1$ with probability $p_i \in [0, 1]$ and $X_i = 0$ otherwise) and set $\mu := \mathrm{Ex}[\sum_{i=1}^{n} X_i]$. Then, for every $0 < \epsilon \leq 1$,*

$$\Pr\left[\sum_{i=1}^{n} X_i \notin [(1-\epsilon) \cdot \mu, (1+\epsilon) \cdot \mu]\right] \leq \frac{2}{e^{\frac{\epsilon^2}{3R} \cdot \mu}}$$

With this tool at hand we can show that all cuts exceeding a certain threshold in size are preserved. This threshold scales inversely with the reciprocal sampling probability, which suggest that a large sampling probability, that also leads to a dense sparsifier, would be beneficial in terms of obtaining correctness for as many cuts as possible. For technical reasons, we directly proceed with analyzing a more general sampling process in which each edge e is sampled with some individual probability p_e.

```
E_H ← ∅
foreach e ∈ E do
    With probability p: E_H ← E_H ∪ {e} and w_H(e) ← 1/p_e
return H = (V, E_H)
```

Lemma 1.4. *Consider an unweighted graph $G = (V, E)$ with $|V| \leq n$ many nodes (for some $n \geq 1$), and let $p \in [0, 1]$ and $t = 3(c+1)p^{-1}\epsilon^{-2} \ln n$. If $H = (V, E_H)$ is obtained from a graph G by the generalized sampling procedure described above with $p_e \geq p$ for each edge $e \in E$, then for each cut $(S, V \setminus S)$ with $|E_G(S, V \setminus S)| \geq t$ the following inequality holds:*

$$\Pr\left[w_H(E_H(S, V \setminus S)) \notin [(1-\epsilon)|E_G(S, V \setminus S)|, (1+\epsilon)|E_G(S, V \setminus S)|]\right] \leq \frac{1}{n^c}\,.$$

Proof. For every edge $e \in E$, let X_e be the binary random variable that has value $w_H(e) = \frac{1}{p_e}$ if e sampled and 0 otherwise. Thus $w_H(E_H(S, V \setminus S)) = \sum_{e \in E_G(S, V \setminus S)} X_e$ and as further

$$\mu := \mathrm{Ex}[\sum_{e \in E_G(S, V \setminus S)} X_e] = \sum_{e \in E_G(S, V \setminus S)} \mathrm{Ex}[X_e] = \sum_{e \in E_G(S, V \setminus S)} \frac{1}{p_e} \cdot p_e = \sum_{e \in E_G(S, V \setminus S)} 1 = |E_G(S, V \setminus S)| \geq t\,,$$

we have

$$\Pr\left[w_H(E_H(S, V \setminus S)) \notin [(1-\epsilon)|E_G(S, V \setminus S)|, (1+\epsilon)|E_G(S, V \setminus S)|]\right] = \Pr\left[\sum_{e \in E_G(S, V \setminus S)} X_e \notin [(1-\epsilon)\mu, (1+\epsilon)\mu]\right].$$

As $p_e \geq p$ for every $e \in E$, we have $X_e \leq R$ for $R := \frac{1}{p}$ and thus by the Chernoff bound above the failure probability for cut $(S, V \setminus S)$ is at most

$$\frac{2}{e^{\frac{\epsilon^2}{3R} \cdot \mu}} \leq \frac{2}{e^{\frac{\epsilon^2 p}{3} \cdot t}} = \frac{2}{e^{(c+1) \ln n}} = \frac{2}{n^{c+1}} \leq \frac{1}{n^c}\,,$$

where the last inequality follows from the implicit assumption $|V| \geq 2$ (that we can safely make because for $|V| = 1$ the edge sampling process has no effect). □

Inspecting this proof, we observe that we can actually get an even more interesting guarantee in which the probability of failure decreases with the cut size.

Lemma 1.5. *Let $p \in [0,1]$ and $t = 3(c+1)p^{-1}\epsilon^{-2}\ln n$. If H is obtained from G by the sampling procedure described above with $p_e \geq p$ for each edge $e \in E$, then for each cut $(S, V \setminus S)$ with $|E_G(S, V \setminus S)| = \alpha t$ for some $\alpha \geq 1$ the following inequality holds:*

$$\Pr\left[w_H(E_H(S, V \setminus S)) \notin [(1-\epsilon)|E_G(S, V \setminus S)|, (1+\epsilon)|E_G(S, V \setminus S)|]\right] \leq \frac{1}{n^{c\alpha}}\,.$$

1.2 Cut Sparsification with Large Minimum Cut Size

To leverage this decrease in the failure probability with increasing cut size, we need a polynomial bound on the number of cuts of a certain size. We can get such a bound relative to the size of a minimum cut.

Lemma 1.6. *Let λ_G denote the size of a minimum cut of a graph G. Then, for every $\alpha \geq 1$, the number of cuts $(S, V \setminus S)$ with $|E_G(S, V \setminus S)| = \alpha\lambda_G$ is at most $2n^{2\alpha}$.*

Proof. The statement follows from the analysis of Karger's contraction algorithm. □

Now a union bound gives us a first sparsification result for graphs of large enough minimum cut.

Lemma 1.7. *Let $p \in [0,1]$ and $t = 3(c+5)p^{-1}\epsilon^{-2}\ln n$. If the minimum cut of G has size $\lambda_G \geq t$ and H is obtained from G by the sampling procedure described above with $p_e \geq p$ for each edge $e \in E$, then H is a $(1 \pm \epsilon)$-cut sparsifier of G with probability at least $1 - \frac{1}{n^c}$.*

Proof. As G is an unweighted graph, every cut has an integral size in the range $\{\lambda_G, \ldots, n\}$ (where $\lambda_G \geq t > 0$). For each $i \in \{\lambda_G, \ldots, n\}$, set $\alpha_i = \frac{i}{\lambda_G}$. Furthermore, for each $i \in \{\lambda_G, \ldots, n\}$, let $(S_1^{(i)}, V \setminus S_1^{(i)})$, ..., $(S_{k_i}^{(i)}, V \setminus S_{k_i}^{(i)})$ denote all cuts of G of size $i = \alpha_i \cdot \lambda_G$, where by Lemma 1.6 we know that $k_i \leq 2n^{2\alpha_i}$. Now for every $i \in \{\lambda_G, \ldots, n\}$ and every $j \in \{1, \ldots, k_i\}$), let $A_j^{(i)}$ be the event "$w_H(E_H(S_j^{(i)}, V \setminus S_j^{(i)})) \notin [(1-\epsilon)|E_G(S_j^{(i)}, V \setminus S_j^{(i)})|, (1+\epsilon)|E_G(S_j^{(i)}, V \setminus S_j^{(i)})|$". Note that H is a $(1 \pm \epsilon)$-cut sparsifier if and only if the event $\bigcup_{i=\lambda_G}^{n} \bigcup_{j=1}^{k_i} A_j^{(i)}$ does not occur. For each $i \in \{\lambda_G, \ldots, n\}$ and $j \in \{1, \ldots, k_i\}$), we have $\Pr[A_j^{(i)}] \leq \frac{1}{n^{(c+4)\alpha_i}}$. By the union bound and Lemma 1.6 we obtain

$$\Pr\left[\bigcup_{i=\lambda_G}^{n} \bigcup_{j=1}^{k_i} A_j^{(i)}\right] \leq \sum_{i=\lambda_G}^{n} \sum_{j=1}^{k_i} \Pr[A_j^{(i)}] \leq \sum_{i=\lambda_G}^{n} \sum_{j=1}^{k_i} \frac{1}{n^{(c+4)\alpha_i}} = \sum_{i=1}^{n} \frac{k_i}{n^{(c+4)\alpha_i}} \leq \sum_{i=1}^{n} \frac{2n^{2\alpha_i}}{n^{(c+4)\alpha_i}}$$
$$= \frac{2n^{2\alpha_i+1}}{n^{(c+4)\alpha_i}} \leq \frac{1}{n^c}\,.$$

As explained above, this means that H is a H is a $(1 \pm \epsilon)$-cut sparsifier with probability at least $1 - \frac{1}{n^c}$. □

Consider for a moment the case the setting that each edge is sampled with the same probability p. Then the required minimum cut size is $\Omega(p^{-1}\epsilon^{-2}\ln n)$ whereas the expected number of edges of H is pm. If we want to keep the requirement on the minimum cut size as mild as possible, we would set p to a constant, like for example $p = \frac{1}{2}$. The resulting guarantee of $pm = m/2$ on the expected sparsifier size sounds a bit weak at first sight. But once we get rid of the requirement on the minimum cut size, such a constant-factor decrease is actually very desirable because then we can sparsify the sparsifier again to obtain an even sparser sparsifier.

1.3 The Half-Sparsify Algorithm

The basic idea to also deal with graphs of small minimum cut size is that we can directly add a small subgraph containing all small cuts to the sparsifier, and then after removing all small cuts the minimum cut size is small and we can proceed as above. Formally, this small subgraph is defined as follows.

Definition 1.8. *A t-bundle $B = (V, E_1 \cup \cdots, E_t)$ (for some $t \geq 1$) is a subgraph of $G = (V, E)$ such that, for all $1 \leq i \leq t$, (V, E_i) is a spanning forest of $(V, E_1 \cup \cdots, E_{i-1})$.*

Note that a straightforward way of computing a t-bundle is to repeatedly (a) compute a spanning forest of the graph and (b) remove it from the graph.

It is a straightforward observation that t-bundles preserve all cuts up to size t.

Observation 1.9. *Let $B = (V, E_B)$ be a t-bundle of $G = (V, E)$. Then, for every cut $(S, V \setminus S)$, we have $|E_G(S, V \setminus S) \cap E_B| = \min(t, |E_G(S, V \setminus S)|)$.*

Our "half-sparsify" algorithm for computing a cut sparsifier with the guarantee explained above now proceeds as follows.

Set $p = \frac{1}{4}$ and $t = 3(c+6)p^{-1}\epsilon^{-2}\ln n$
Let $B = (V, E_B)$ be a $\lfloor t \rfloor$-bundle of G
$E_{G'} \leftarrow \emptyset$
foreach $(u, v) \in E$ **do**
  With probability p: $E_{G'} \leftarrow E_{G'} \cup \{e\}$ and $w_{G'}(e) \leftarrow \frac{1}{p}$
return $H = (V, E_B \cup E_{G'})$

To proceed with the proof, we need the following decomposition lemma whose proof is a simple exercise. It essentially states that sparsifiers of subgraphs can be combined to obtain a sparsifier of the full graph.

Lemma 1.10 (Decomposability). *Let $H_1 = (V, E_{H_1})$ be a $(1 \pm \epsilon)$-cut sparsifier of a graph $G_1 = (V, E_{G_1})$ and let $H_2 = (V, E_{H_2})$ be a $(1 \pm \epsilon)$-cut sparsifier of a graph $G_2 = (V, E_{G_2})$. Show that $H = (V, E_{H_1} \cup E_{H_2})$ is a $(1 \pm \epsilon)$-cut sparsifier of $G = (V, E_{G_1} \cup E_{G_2})$.*

Lemma 1.11. *If H is obtained from G by the half-sparsify algorithm above, then, with probability at least $1 - \frac{1}{n^c}$, $H = B \cup G'$ is a $(1 \pm \epsilon)$-cut sparsifier of G with $|E_B| = O(cn\epsilon^{-2} \log n)$ and $|E_{G'}| \leq \frac{|E_G|}{2}$.*

Proof. The size analysis is straightforward. Each spanning forest of the $\lfloor t \rfloor$-bundle B contributes at most $n - 1$ edges, and thus $|E_B| = O(tn)$. Furthermore $\mathrm{Ex}[|E_{G'}|] \leq \frac{|E_G|}{4}$ and by applying the Chernoff bound we get that $|E_{G'}| \leq \frac{|E_G|}{2}$ with probability at least $1 - \frac{1}{n^c}$.

To analyze the correctness, let $C \subseteq V$ be the set of nodes of an arbitrary connected component of $G \setminus B$.[1] Let us compare $G[C]$, the subgraph of G (the original graph) induced by C, with $(G \setminus B)[C]$,

[1] It might be the case that $C = V$; the proof then becomes trivial, but we avoid a case distinction.

the connected component of $G \setminus B$. Since $(G \setminus B)[C]$ is connected its minimum cut size is at least 1. Thus, removing the edges of the $\lfloor t \rfloor$-bundle has not reduced the size of any cut to 0, which, by Observation 1.9, implies that $G[C]$ has a minimum cut size of at least $\lfloor t \rfloor + 1 \geq t$.

Note that the construction of $G[C]$ (which is determined by the $\lfloor t \rfloor$-bundle) is independent of the random choices of our algorithm. We can now alternatively view $H[C]$, the subgraph of H (the returned graph) induced by C, as the result of a sampling process in which we set a sampling probability of $p_e = 1$ to every edge $e \in E_B$ and a sampling probability of $p_e = p$ to every edge $e \in E \setminus E_B$. It is then still true that construction $G[C]$ is independent of the (actual) random choices of this sampling process. Therefore the probabilistic guarantee of Lemma 1.7 applies to $G[C]$. As argued above, the minimum cut size of $G[C]$ is $\lambda_{G[C]} \geq t$, and thus $H[C]$ is a $(1 \pm \epsilon)$-cut sparsifier of $G[C]$ with probability at least $1 - \frac{1}{n^{c+1}}$.

Finally, let $C_1, \ldots, C_k$ denote the sets of nodes of the connected components of $G \setminus B$ and let $B' = (V, E_{B'})$ be the subgraph of G containing all edges with endpoints in different sets, i.e., $E_{B'} = \{(u, v) \in E \mid u \in C_i, v \in C_j, i \neq j\}$. Then, as argued above, $H[C_i]$ is a $(1 \pm \epsilon)$-cut sparsifier of $G[C_i]$ for each $1 \leq i \leq k$. Furthermore, B' is trivially a $(1 \pm \epsilon)$-cut sparsifier of B' itself. As $H = H[C_1] \cup \cdots \cup H[C_k] \cup B'$, we get by Lemma 1.10 and the union bound that H is a $(1 \pm \epsilon)$-cut sparsifier of G with probability at least $1 - \frac{1}{n^c}$. □