

Lecture 13: Graph Sparsification

Notes by Michael Kapralov¹

In this lecture we show how to construct *cut sparsifiers* of graphs. More formally, we show how given an unweighted graph $G = (V, E)$, one can construct a weighted graph $G' = (V, E, w)$ with $O(\frac{1}{\epsilon^2} n \log n)$ edges such that for every $S \subseteq V$ one has

$$(1 - \epsilon)E(S, V \setminus S) \leq E'(S, V \setminus S) \leq (1 + \epsilon)E(S, V \setminus S)$$

with high probability. Here $E(S, V \setminus S)$ stands for the number of edges going from S to $V \setminus S$ in G , and $E'(S, V \setminus S)$ stands for the total weight of edges going from S to $V \setminus S$ in G' .

1 Warm-up: sparsifying the clique

To see that this is even possible let us consider the simplest possible example, i.e. when the graph G is a clique. Let $p = \frac{C \log n}{\epsilon^2 n}$ for a sufficiently large absolute constant $C > 0$. For every edge $e \in E$ include e into E' independently with probability p , giving it weight $1/p$ if it is included. Why does this produce a cut sparsifier with high probability? First, observe that cut sizes are exactly correct in expectation. Indeed, for every $S \subseteq V$ one has

$$\mathbb{E}[E'(S, V \setminus S)] = \sum_{e \in \delta(S)} \frac{1}{p} \cdot \Pr[e \in E'] = E(S, V \setminus S),$$

where $\delta(S)$ stands for the set of edges in E that have one endpoint in S . We can use the Chernoff bound to show that for every $S \subseteq V$ one has

$$(1 - \epsilon)E(S, V \setminus S) \leq E'(S, V \setminus S) \leq (1 + \epsilon)E(S, V \setminus S) \quad (*)$$

with high probability.

Theorem 1 (*Chernoff Bound*) let $Y = \sum_{i=1}^n X_i$ where X_i 's are independent bernoulli random variables, such that $X_i = 1$ with probability p_i and 0 otherwise. Also, let μ_Y denote $\mathbb{E}(Y)$. Then for any $\delta \in (0, 1)$,

$$\Pr[|Y - \mu_Y| > \delta \mu_Y] \leq 2e^{-\frac{\delta^2 \mu_Y}{3}}$$

Fix $S \subset V$, and let $Y = \sum_{e \in \delta(S)} \mathbb{I}[e \in E']$. We have $\mu_Y = \mathbb{E}[Y] = p \cdot E(S, V \setminus S) = p|S|(|V| - |S|) = pr(n-r)$, where we let $r = |S|$ for convenience. Note that it's enough to consider $r \leq n/2$. By Theorem 1 we have

$$\begin{aligned} \Pr[|Y - \mu_Y| > \epsilon \mu_Y] &\leq 2e^{-\frac{\epsilon^2 \mu_Y}{3}} \\ &= 2e^{-\epsilon^2 pr(n-r)/3} \\ &\leq 2e^{-(C/6)r \log n}, \end{aligned}$$

where in the last transition we lower bounded $n - r$ by $n/2$ and used the definition of p . We now take a union bound over all cuts $(S, V \setminus S)$ with $r \leq n/2$ vertices on the smaller side, getting

$$\Pr[\exists S \subseteq V, |S| = r \text{ such that } (*) \text{ fails for } S] \leq \binom{n}{r} \cdot 2e^{-(C/6)r \log n} \leq n^r \cdot 2e^{-(C/6)r \log n} = 2n^{-(C/6-1)r}.$$

¹**Disclaimer:** These notes were written as notes for the lecturer. They have not been peer-reviewed and may contain inconsistent notation, typos, and omit citations of relevant works.

A union bound over all $r = 1, 2, \dots, n/2$ then gives that (*) is satisfied for all S with probability at least

$$1 - \sum_{r=1}^{n/2} 2n^{-(C/6-1)r} \geq 1 - \sum_{r=1}^{\infty} 2n^{-(C/6-1)r} \leq 1 - n^{-1}$$

as long as C is a sufficiently large absolute constant.

2 Uniform sampling

The argument above only applies to the clique, as it is easy to get a good upper bound on the number of cuts of a given size in a clique. A seminal result of Karger [1] shows how to generalize this argument to graphs with a large min-cut. Suppose that G has min-cut k (think of k as large). We now upper bound the number of cuts of size αk for $\alpha \geq 1$ such that 2α is an integer. Run Karger's randomized contraction process for $n - 2\alpha$ steps, and then choose a uniformly random subset of the remaining 2α vertices that contains vertex 1. The probability that a given cut of size αk survives the contraction process is

$$\left(1 - \frac{2\alpha}{n}\right) \cdot \left(1 - \frac{2\alpha}{n-1}\right) \cdots \left(1 - \frac{2\alpha}{2\alpha+1}\right) = \left(\frac{n}{2\alpha}\right)^{-1},$$

and the probability that the cut is output conditioned on having survived the contraction process is exactly equal to $2^{1-2\alpha}$. Thus, the number of cuts of size bounded by αk is at most

$$\binom{n}{2\alpha} 2^{2\alpha-1}.$$

Now to construct a sparsifier of G we include every edge $e \in E$ into E' independently with probability $p = \frac{C \log n}{\epsilon^2 k}$ for a sufficiently large constant $C > 0$. We can now take a union bound over cuts of all sizes, getting

$$\begin{aligned} \Pr[\exists S \subseteq V : (*) \text{ fails for } (S, V \setminus S)] &\leq \sum_{\alpha=2}^{\infty} \Pr[\text{a fixed cut of size } \geq (\alpha-1)k \text{ fails}] \cdot \#(\text{cuts of size } \leq \alpha k) \\ &\leq \sum_{\alpha=2}^{\infty} 2e^{-\epsilon^2(\alpha-1)kp/3} \cdot \binom{n}{2\alpha} 2^{2\alpha-1} \\ &\leq 2 \sum_{\alpha=2}^{\infty} e^{-\epsilon^2(\alpha-1)kp/3} \cdot (2n)^{2\alpha} \\ &\leq 2 \sum_{\alpha=2}^{\infty} n^{-(C/3)(\alpha-1)} \cdot (2n)^{2\alpha} \\ &\leq 2 \sum_{\alpha=2}^{\infty} n^{-(C/6)\alpha} \cdot (2n)^{2\alpha} \\ &\leq 1/n \end{aligned}$$

as long as C is a sufficiently large absolute constant.

Exercise 1 Prove that if G is a bipartite clique and $p = o(\frac{\log n}{n})$, then G' is not a sparsifier of the clique on n vertices with $\epsilon = 1/2$ with high probability.

3 Non-uniform sampling

In the previous section we showed that uniform sampling can be used to obtain a sparsifier with few edges when the min-cut in the graph is large. If the min-cut is small, uniform sampling will not succeed. Indeed, suppose that the input graph G is a union of two disjoint cliques of size $n/2$ joined by a single edge: any sparsifier must include that edge, but uniform sampling will either lose it or not reduce the number of edges in the cliques. The right approach here is to keep the edge connecting the cliques and sampled edges induced by the cliques, giving sampled edges appropriately large weight, as we did in the previous section. To generalize this idea, we need a convenience definition of ‘well-connected’ parts of the graph. This is provided by

Definition 2 (Strong connectivity) *A k -strongly connected component of a graph $G = (V, E)$ is a maximal induced subgraph that is k -edge connected (i.e., with min-cut at least k). For every edge $e \in E$ the strong connectivity of k_e of e is the maximum k such that both endpoints of e belong to a k -strongly connected component.*

Exercise 2 *Prove that for every $k_1 \leq k_2$ and every graph $G = (V, E)$ the collection of k_2 -strongly connected components in G is a refinement of the collection of k_1 -strongly connected components in G .*

We will let $p_e = \min \left\{ \frac{r}{k_e}, 1 \right\}$, where $r = \frac{C \log n}{\epsilon^2}$ for a sufficiently large constant $C > 0$. Thus, the expected number of edges in the sparsifier will be bounded by

$$\sum_{e \in E} p_e = \sum_{e \in E} \min \left\{ \frac{r}{k_e}, 1 \right\} \leq \sum_{e \in E} \frac{r}{k_e} = r \sum_{e \in E} 1/k_e.$$

To bound its size, we prove

Lemma 3 $\sum_{e \in E} 1/k_e \leq n - 1$.

Proof The proof is by induction. The base case corresponds to $n = 1$, in which case the claim trivially holds. Now consider a graph G and a minimum cut $(S, V \setminus S)$ in G . Let k denote the number of edges in the min cut. The strong connectivity of every edge in the min-cut is at least k , since G is k -connected. Thus, the total contribution of edges crossing the cut to the sum is at most $\sum_{e \in \delta(S)} 1/k_e \leq \sum_{e \in \delta(S)} 1/k = k \cdot 1/k = 1$. On the other hand, we have by the inductive hypothesis

$$\sum_{e \in E, e \text{ induced by } S} 1/k_e^S \leq |S| - 1 \text{ and } \sum_{e \in E, e \text{ induced by } V \setminus S} 1/k_e^{V \setminus S} \leq |V \setminus S| - 1,$$

where k_e^S and $k_e^{V \setminus S}$ denote strong connectivities in the subgraphs induced by S and $V \setminus S$ respectively. It remains to note that for all e induced by S one has $k_e^S \leq k_e$, and similarly for all e induced by $V \setminus S$ one has $k_e^{V \setminus S} \leq k_e$, so the above imply

$$\sum_{e \in E, e \text{ induced by } S} 1/k_e \leq |S| - 1 \text{ and } \sum_{e \in E, e \text{ induced by } V \setminus S} 1/k_e \leq |V \setminus S| - 1.$$

Putting everything together, we get

$$\sum_{e \in E} 1/k_e \leq (|S| - 1) + (|V \setminus S| - 1) + 1 \leq |V| - 1,$$

as required. ■

We now prove that if G' is obtained by including every edge $e \in E$ independently with probability p_e (with weight $1/p_e$ if included), then G' is a sparsifier of G with high probability, proving the result

of [3]. The main challenge is that some edges in G' will have very large weights. This means that the random variables corresponding to cut sizes have large variance, and therefore the analysis of the uniform sampling scheme from the previous lecture does not go through. As we will shortly see, this difficulty can be circumvented by a nice decomposition of the graph into layers corresponding to components of different connectivity.

Consider a reweighted graph G_w , where edges are given weight k_e (assume for simplicity that all edges have $p_e < 1$). Thus, our sampling process is simply the process of including every edge in G_w independently with probability p_e with its own weight. Let $k_1 < k_2 < \dots < k_T$ denote distinct values of edge strengths in G (note that $T \leq m$, where m is the number of edges). For every $t = 1, \dots, T$ let $F_{\geq t}$ denote the set of edges in G whose strength is at least k_t , and note that

$$G_w = \sum_{t=1}^T (k_t - k_{t-1}) F_{\geq t},$$

where $k_0 = 0$ by convention. To see that the above equality holds, note that the weight of e on the lhs is k_e , and its weight on the rhs is $\sum_{t=1}^{t'} (k_t - k_{t-1}) = k_e$, where t' is the largest such that $e \in F_{\geq t'}$. Now consider a sampling process where if an edge $e \in E$ is included in G (with probability p_e), then it is included in all $F_{\geq t}$ that it belongs to, namely in all $F_{\geq t}$ for $t \leq k_e$. The sampling processes in $F_{\geq t}$ are correlated, but the sampling within any given $F_{\geq t}$ is independent, so it is enough to prove that all cuts in all $F_{\geq t}$'s concentrate around their expectation. In order to show that, we use the following stronger version of the uniform sampling claim proved in the previous section:

Theorem 4 [2] *For a graph $G = (V, E)$, if every edge $e \in E$ is included independently with probability p_e , and the expected number of edges crossing every cut is at least $(C \log n)/\epsilon^2$ for a sufficiently large constant $C > 0$, then in the resulting graph all cuts are within a $1 \pm \epsilon$ factor of their expectation with probability at least $1 - 1/n$.*

Its proof is an exercise, given the analysis of our uniform sampling scheme from the previous section. Note that we will be applying Theorem 4 to the $F_{\geq t}$'s, which is exactly what lets us circumvent the issue of large weights. Now it remains to note that expected sizes of cuts in $F_{\geq t}$ are at least r . To verify that, consider any connected component in $F_{\geq t}$, and any cut in this component. Suppose that the number of edges of G crossing this cut is k' . Then the strong connectivity of all of those edges is at most k' , so the probability of sampling those edges is at least r/k' , and therefore the expected number of edges in the cut is at least $r = (C \log n)/\epsilon^2$. Thus, the preconditions of Theorem 4 are satisfied, and we get that G' is a sparsifier of G with high probability.

References

- [1] David R. Karger: Global Min-cuts in RNC, and Other Ramifications of a Simple Min-Cut Algorithm, SODA'93
- [2] David R. Karger: Random sampling in cut, flow, and network design problems, STOC'94
- [3] András A. Benczúr and David R. Karger: Randomized Approximation Schemes for Cuts and Flows in Capacitated Graphs, SIAM Journal of Computing, 2015