1   Graph Sparsifiers

Let $G = (V, E)$ be an undirected graph. How well can G be approximated by a sparse graph? Such questions have been studied for various notions of approximation. Today we will look at **approximating the cuts** of the graph. As before, the cut defined by $U \subseteq V$ is

$$\delta(U) = \{uv \in E \ : \ u \in U \text{ and } v \notin U\}.$$

Let $w : E \to \mathbb{R}$ be a weight function on the edges. The **weight** of a set $F \subseteq E$ is defined to be

$$w(F) \ := \ \sum_{e \in F} w_e.$$

So the weight of the cut defined by $U \subseteq V$ is $w(\delta(U))$.

A **graph sparsifier** is a non-negative weight function $w : E \to \mathbb{R}$ such that

- w has only a small number of non-zero weights,
- for every $U \subseteq V$,
$$(1 - \epsilon)|\delta(U)| \ \leq \ w(\delta(U)) \ \leq \ (1 + \epsilon)|\delta(U)|. \tag{1}$$

We can think of any edge with weight zero as being deleted. So the goal is to find a sparse, but weighted, subgraph of G such that the weight of every cut is preserved up to a multiplicative factor of $1 + \epsilon$.
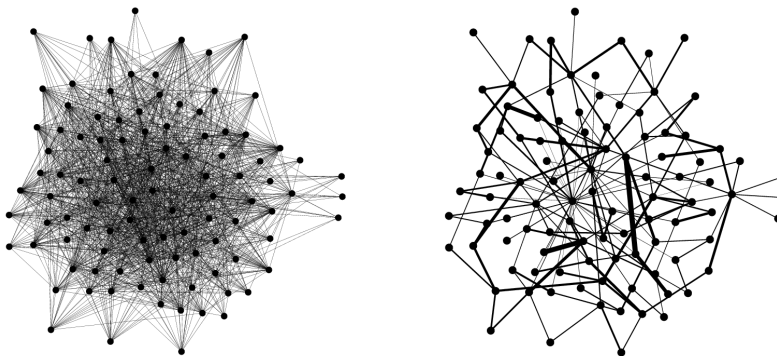
How could one find a sparsifier? A natural idea is to sample the edges independently with some probability p. That works well if G is the complete graph because it essentially amounts to constructing an Erdos-Renyi random graph, which is well-studied.

Unfortunately this approach falls apart when G is quite different from the complete graph. One such graph is the "dumbbell graph", which consists of two disjoint cliques, each on $n/2$ vertices, and a single edge in the middle connecting the cliques. We would like to get rid of most edges in the cliques, but we would need to keep the edge in the middle. This example tells us that we should not sample all edges with the same probability p.
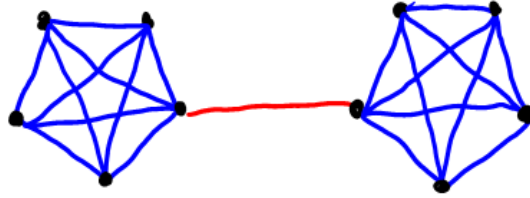
So now the question is: for each edge, how "important" is it? Should we sample it with low probability or high probability? The notion of **edge-connectivity**, which we defined in the previous lecture, seems quite useful. Recall that for an edge e we let k_e be the minimum size of a cut containing e, i.e.,

$$k_e := \min\{ |\delta(U)| : U \subset V \text{ and } e \in \delta(U) \}.$$

By the Max-Flow Min-Cut theorem, k_e equals the maximum amount of flow that can be sent between the endpoints of the edge. So k_e can be efficiently computed. Edges with high connectivity only appear in cuts with many other edges, so intuitively they are not terribly important. In the dumbbell example, the clique edges have connectivity $n/2 - 1$, whereas the single edge in the middle has connectivity 1. So the connectivity values seem to do a good job at identifying important edges.

1.1 The sampling process

Consider the following sampling process, where ρ is a parameter the determines the number of "rounds" of sampling.

- Initially $w = 0$.
- Compute all edge connectivities k_e.
- For $i = 1, \ldots, \rho$
-     For each $e \in E$
-         With probability $1/k_e$, increase w_e by k_e/ρ.

One great feature of this sampling process is that all edge weights are preserved in expectation.

Claim 1 *For every edge e we have* $\mathrm{E}[w_e] = 1$.

PROOF: The expected increase in w_e in the ith iteration is $(1/k_e) \cdot (k_e/\rho) = 1/\rho$. By linearity of expectation, the expected increase over all ρ iterations is 1. □

Moreover, the weight of any set of edges is also preserved in expectation.

Corollary 2 *For any $F \subseteq E$, we have* $\mathrm{E}[w(F)] = |F|$.

PROOF: By linearity of expectation. □

In particular, $\mathrm{E}[w(\delta(U))] = |\delta(U)|$ for every U. Unfortunately preserving cuts in expectation is not good enough. We would like to say that, with high probability, *every* cut's weight is is close to its expectation. This is a statement about concentration, so the Chernoff bound seems like a natural tool to try.

1.2 An analysis that doesn't work

Consider some cut $\delta(U)$. The weight of that cut after sampling is the random variable

$$X \; := \; \sum_{i=1}^{\rho} \sum_{e \in \delta(U)} \frac{k_e}{\rho} X_{i,e},$$

where $X_{i,e}$ is a Bernoulli random variable indicating whether edge e was sampled during the ith round of sampling. The Chernoff bound is designed for analyzing sums of independent Bernoullis, so it seems that we are in great shape.

But there is a problem: the coefficients k_e/ρ. The Chernoff bound works for any sum of independent random variables taking values in $[0, 1]$. Unfortunately we have these coefficients that can be quite large (e.g., $k_e \approx n$), so a straightforward application of Chernoff will not work.

Actually it is not really a problem that the coefficients are *big*, but it is a problem that they could be *wildly different*. Consider the example:

$$X \; = \; \sum_{i=1}^{n} n \cdot X_i$$

where the X_i's are independent Bernoulli random variable. Even though there are these large coefficients, we can still analyze X with the Chernoff bound because it is simply n times the random variable $\sum_i X_i$, which has no coefficients. On the other hand, consider the example:

$$Y \; = \; n \cdot Y_0 + \sum_{i=1}^{n} Y_i$$

where $Y_0, \ldots, Y_n$ are Bernoulli random variables that are 1 with probability $1/n$. Then $\mathrm{E}[Y] = 2$, but $\Pr[Y \geq n] \geq 1/n$. The Chernoff bound cannot directly give useful tail bounds on Y because it is designed to show that the probability of being α times larger than the expectation is *exponentially decreasing* in α, and that is simply not true for Y.

1.3 Analysis trick: group edges by connectivity

To handle the different weights, we will partition the edges into groups whose connectivities are roughly the same. Formally, we set $E = E_1 \cup \cdots \cup E_{\log n}$ where
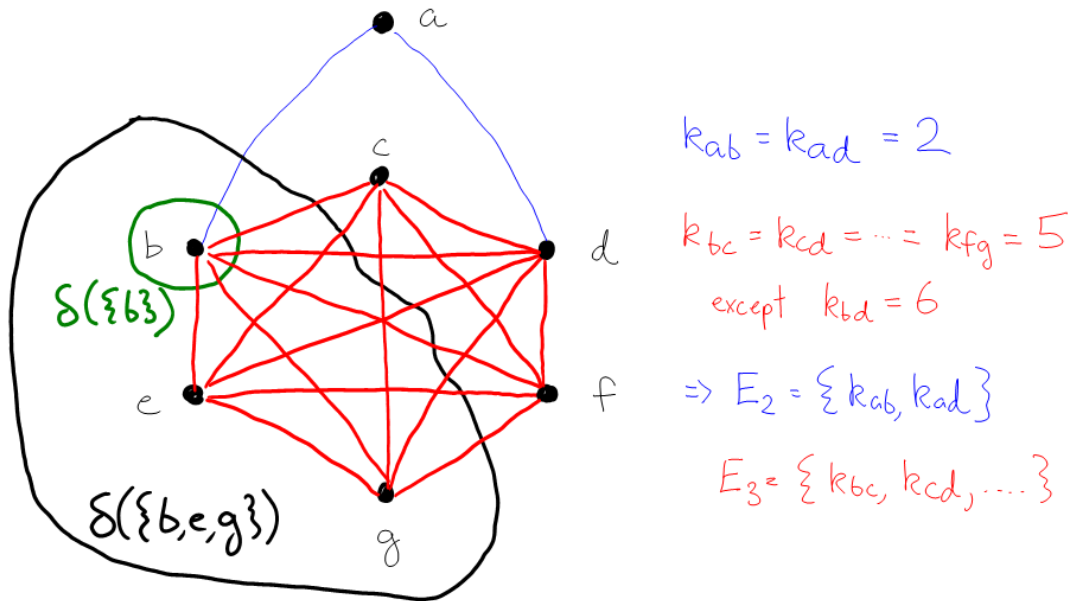
$$E_i \; = \; \{\, e \in E \; : \; 2^{i-1} \leq k_e < 2^i \,\}.$$

Now instead of proving concentration for $w(\delta(U))$, we will prove concentration for $w(\delta(U) \cap E_i)$, because all edges in E_i have nearly the same coefficients k_e/ρ.

So let F be some set of the form $\delta(U) \cap E_i$. We will call such a set a **cut-induced set**. Note that there could be some other cut U' such that $F = \delta(U) \cap E_i = \delta(U') \cap E_i$. We're going to focus on the *smallest* such cut, because we want the sampling error of $w(F)$ to be small relative to $|\delta(U)|$, and that is hardest when $|\delta(U)|$ is small. So define

$$q(F) \; = \; \min\{\, |\delta(U)| \; : \; U \subseteq V \; \wedge \; \delta(U) \cap E_i = F \,\}.$$

This definition is a bit hard to understand, so we illustrate it with the following example. The set $\{ab\}$ is a cut-induced subset of E_2 because $\delta(\{b\}) \cap E_2 = \{ab\}$. But we also have $\delta(\{b, e, g\}) \cap E_2 = \{ab\}$. The smallest cut that induces $\{ab\}$ is $\delta(\{b\})$, so we have $q(\{ab\}) = |\delta(\{b\})| = 6$.

We can use Chernoff bounds to prove the following concentration bound for $w(F)$.

Claim 3 *Let $F \subseteq E_i$ be a cut-induced set. Then*

$$\Pr\left[|w(F) - \mathrm{E}[w(F)]| > \frac{\epsilon q(F)}{\log n} \right] \leq 2\exp\left(-\frac{\epsilon^2 \rho q(F)}{3 \cdot 2^i \log^2 n} \right)$$

The proof is just a calculation, so we save it to Section **??**.

1.4 The Main Theorem

Our main theorem is:

Theorem 4 *Let $G = (V, E)$ be a graph with $n = |V|$. Then with probability at least $1/2$, our sampling process will produce weights $w : E \to \mathbb{R}$ satisfying (1) and with only $O(n \log^3(n)/\epsilon^2)$ non-zero entries.*

By a slightly more careful analysis one can improve the $\log^3 n$ to $\log^2 n$. Instead of edge connectivities, if we use a slightly different quantity to determine the importance of an edge, the $\log^3 n$ can be improved to $\log n$. And by a completely different technique, the $\log^3 n$ can be removed entirely!

To prove our theorem we need the following result which we stated last time, and which follows from a variant of the contraction algorithm.

4

Theorem 5 *Let $G = (V, E)$ be a graph. Let $B \subseteq E$ be arbitrary and let $K \leq \min\{ k_e : e \in B \}$. Then, for every real $\alpha \geq 1$,*

$$|\{ \delta(U) \cap B : U \subseteq V \ \wedge \ |\delta(U)| \leq \alpha K \}| \ < \ n^{2\alpha}.$$

We also need the following fact:

Fact 6 *For any graph $G = (V, E)$ with $n = |V|$ we have $\sum_{e \in E} 1/k_e \leq n - 1$.*

PROOF:(of Theorem 4). We will set $\rho = 100 \log^3(n)/\epsilon^2$.

The number of non-zeros is easy to analyze. Let $X_{i,e}$ be the indicator random variable that is 1 if edge e is sampled in round i, so $\mathrm{E}[X_{i,e}] = 1/k_e$. The number of non-zero weights in w is at most $\sum_{i=1}^{\rho} \sum_{e \in E} X_{i,e}$. So the expected number of non-zero weights is at most $\rho \sum_e 1/k_e = O(n \log^3(n)/\epsilon^2)$, by Fact 6. By Markov's inequality, the number of non-zero weights is at most four times that with probability at least $3/4$.

Now we must show that the cuts are concentrated. We will show that, with high probability, every cut-induced set F satisfies

$$|w(F) - |F|| \ \leq \ \frac{\epsilon q(F)}{\log n}. \tag{2}$$

If this is true, then for every cut $C = \delta(U)$ we have

$$|w(C) - |C|| \ \leq \ \sum_{i=1}^{\log n} |w(C \cap E_i) - |C \cap E_i|| \ \leq \ \sum_{i=1}^{\log n} \frac{\epsilon q(C \cap E_i)}{\log n} \ \leq \ \sum_{i=1}^{\log n} \frac{\epsilon |C|}{\log n} \ = \ \epsilon |C|,$$

where the first inequality comes from the triangle inequality, the second comes from (2), and the third follows since $q(C \cap E_i)$ is the size of the smallest cut that induces $C \cap E_i$. This proves our desired inequality (1).

So it remains to show (2). Fix any $i \in \{1, \ldots, \log n\}$. Let $F^1, F^2, \ldots$ be all the cut-induced subsets of E_i, ordered such that $q(F^1) \leq q(F^2) \leq \ldots$. Define

$$p_j \ := \ \Pr\left[|w(F^j) - \mathrm{E}[w(F^j)]| > \frac{\epsilon q(F^j)}{\log n} \right] \ \leq \ 2\exp\left(-\frac{\epsilon^2 \rho q(F^j)}{3 \cdot 2^i \log^2 n} \right), \tag{3}$$

by Claim 3. We will use a union bound to show that all of the F^j's are concentrated. The trouble is that there can be exponentially many of them, so the argument needs to be a bit clever.

Consider the first n^2 cut-induced sets $F^1, \ldots, F^{n^2}$. Since every edge $e \in F^j$ belongs to E_i we have $k_e \geq 2^{i-1}$. This means that every cut containing that edge has size at least 2^{i-1}, and therefore $q(F^j) \geq 2^{i-1}$. Plugging this into (3), we get

$$p_j \ \leq \ 2\exp\left(-\frac{\epsilon^2(100 \log^3(n)/\epsilon^2)2^{i-1}}{3 \cdot 2^i \log^2 n} \right) \ \leq \ 2\exp(-16 \log n) \ \leq \ 2n^{-16}.$$

So $\sum_{j=1}^{n^2} p_j \leq 2n^{-14}$.

Now we consider the remaining cut-induced sets F^j for $j > n^2$. We now apply Theorem 5 with $B = E_i$, which means we can take $K = 2^{i-1}$. The theorem states that, for any $\alpha \geq 1$,

$$|\{ \text{cut induced set } F \subseteq E_i : q(F) \leq \alpha 2^{i-1} \}| \ < \ n^{2\alpha}.$$

So for any $\alpha \geq 1$, the number of F^j with $q(F^j) \leq \alpha 2^{i-1}$ is less than $n^{2\alpha}$. By our ordering of the F^j's, we must have $q(F^{n^{2\alpha}}) > \alpha 2^{i-1}$. Substituting $\alpha = \frac{\ln j}{2\ln n}$ we get $q(F^j) > \frac{\ln j}{2\ln n} 2^{i-1}$. Plugging into (3) we get

$$p_j \;\leq\; 2\exp\left(-\frac{\epsilon^2(100\log^3(n)/\epsilon^2)\ln(j)2^{i-1}}{6\cdot 2^i\log^2(n)\ln(n)}\right) \;<\; j^{-8}.$$

Thus, summing over all $j > n^2$,

$$\sum_{j>n^2} p_j \;\leq\; \sum_{j>n^2} j^{-8} \;\leq\; \int_{n^2}^{\infty} j^{-8}\,dj \;=\; \left.-\frac{j^{-7}}{7}\right|_{j=n^2}^{\infty} \;<\; n^{-14}.$$

So, by a union bound, the probability that any cut-induced subset of E_i violates (2) is at most

$$\sum_j p_j \;=\; \sum_{j\leq n^2} p_j + \sum_{j>n^2} p_j \;<\; 2n^{-14} + n^{-14} \;<\; n^{-2}.$$

That analysis is for a particular i. Applying a union bound over all $i \in \{1,\ldots,\log n\}$, the total failure probability is at most $1/n$. $\square$