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

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

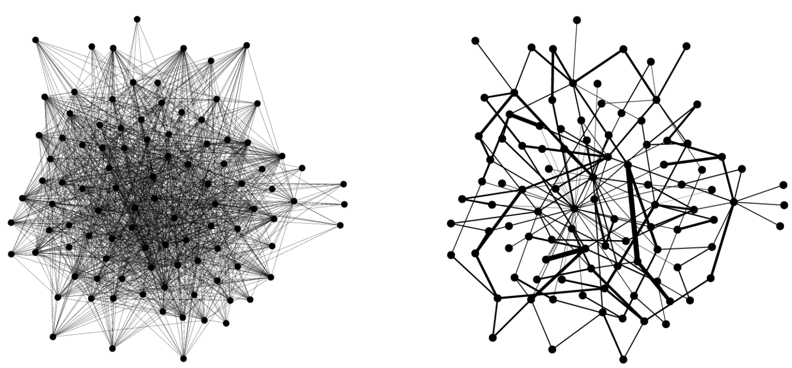
So the weight of the cut defined by $U \subseteq V$ is $w(\delta(U))$.

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 $U \subseteq V$,

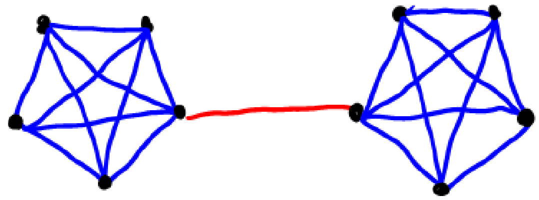
$$(1 - \epsilon)|\delta(U)| \;\; \leq \;\; w(\delta(U)) \;\; \leq \;\; (1 + \epsilon)|\delta(U)|. \tag{1}$$

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 algorithm which independently samples each edge with probability inversely proportional to its connectivity.

Algorithm 1: Algorithm for graph sparsification using edge connectivities. The parameter ρ determines the number of "rounds" of sampling.

```
Initially w_e = 0 for every e ∈ E.
Compute the edge connectivity k_e for every e ∈ E.
for i = 1, ..., ρ do
    foreach e ∈ E do
        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 i^{th} 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 idea: connectivity classes and projections

Often when analyzing quantities with very different magnitudes, it is useful to divide them into groups whose values are roughly the same. We will use this idea to group the edges of the graph, where the quantity of interest is the edge-connectivity. Formally, we set $E = E_1 \cup \cdots \cup E_{\log n}$ where

$$E_i \;=\; \left\{\, e \in E \,:\, 2^{i-1} \leq k_e < 2^i \,\right\}. \tag{2}$$

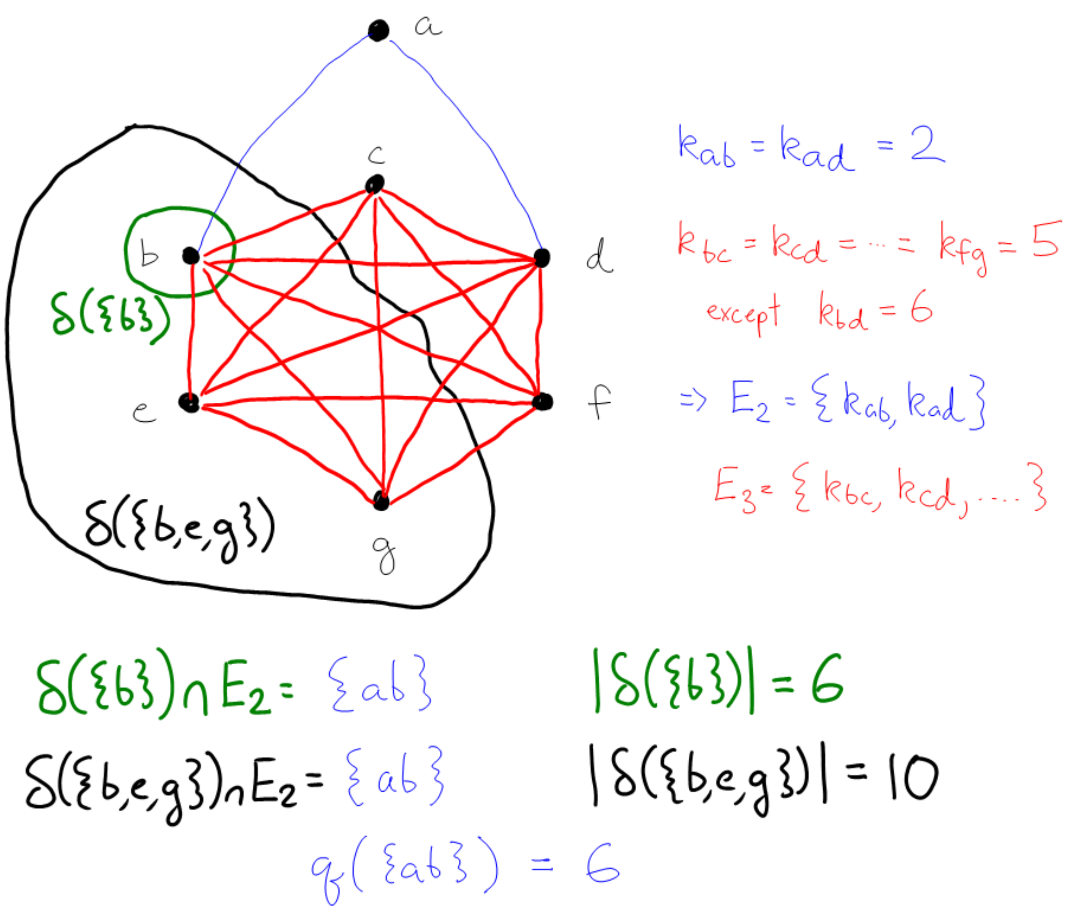
Each set E_i is called a ***connectivity class***.

Instead of trying to show that the weight $w(\delta(U))$ of each cut is concentrated, we will restrict our attention to edges in the same connectivity class. We will show that $w(\delta(U) \cap E_i)$ is concentrated, for each $U \subseteq V$ and each i. Applying a Chernoff bound to $w(\delta(U) \cap E_i)$ will work very nicely because all edges in E_i have nearly the same coefficients k_e/ρ.

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

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

This definition is a bit hard to understand, so we illustrate it with the following example. The set $\{ab\}$ is a projection in 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 projects onto $\{ab\}$ is $\delta(\{b\})$, so we have $\mathrm{sm}(\{ab\}) = |\delta(\{b\})| = 6$.



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

Claim 3 *Let $P \subseteq E_i$ be a projection of a cut. Then*

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

The proof is just a calculation, so we skip it for now.

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 non-random techniques, 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.

Theorem 5 *Let $G = (V, E)$ be a graph. Let E_i be a connectivity class. For every real $\alpha \geq 1$,*

$$|\{\ \delta(U) \cap E_i\ :\ U \subseteq V\ \ \wedge\ \ |\delta(U)| \leq \alpha 2^{i-1}\ \}|\ <\ n^{2\alpha}.$$

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 $\rho = 24 \log^3(n)/\epsilon^2$.

Sparsity analysis. 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, there is probability at most $1/4$ that the number of non-zero weights exceeds its expectation by a factor of four.

Error of Projections. For any projection P, define the event

$$\mathcal{E}_P\ =\ \left\{ |w(P) - \mathrm{E}[\,w(P)\,]|\ >\ \frac{\epsilon \cdot \mathrm{sm}(P)}{\log n} \right\}.$$

Suppose that none of the events $\mathcal{E}_P$ holds. Then for every cut $C = \delta(U)$ we have

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

The first inequality follows from the triangle inequality. The second inequality follows because we have assumed that $\mathcal{E}_{C \cap E_i}$ doesn't hold. The third inequalty holds since $\mathrm{sm}(C \cap E_i)$ is the size of the *smallest* cut whose projection onto E_i is $C \cap E_i$, so $\mathrm{sm}(C \cap E_i) \leq |C|$. This proves our desired inequality (1).

Concentration of Projections. It remains to analyze the probability of avoiding the events $\mathcal{E}_P$. Fix any $i \in \{1, \ldots, \log n\}$. Let $\mathcal{P}_i$ be the collection of *all* projections in E_i. Note that every $P \in \mathcal{P}_i$ contains only edges of connectivity at least 2^{i-1}, so it can only be the projection of a cut of size at least 2^{i-1}. In other words,

$$\mathrm{sm}(P) \geq 2^{i-1} \qquad \forall P \in \mathcal{P}_i. \tag{3}$$

We use a tricky union bound to analyze the probability of avoiding the events $\mathcal{E}_P$.

$$\begin{aligned}
\Pr\left[\bigcup_{P\in\mathcal{P}_i}\mathcal{E}_P\right] &\leq \sum_{P\in\mathcal{P}_i}\Pr\left[\,\mathcal{E}_P\,\right] \qquad \text{(union bound)}\\
&= \sum_{j\geq i}\sum_{\substack{P\in\mathcal{P}_i\\ 2^{j-1}\leq \mathrm{sm}(P)<2^j}}\Pr\left[\,\mathcal{E}_P\,\right] \qquad \text{(by (3))}\\
&\leq \sum_{j\geq i}\sum_{\substack{P\in\mathcal{P}_i\\ 2^{j-1}\leq \mathrm{sm}(P)}} 2\exp\left(-\frac{\epsilon^2\rho\cdot\mathrm{sm}(P)}{3\cdot 2^i\log^2 n}\right) \qquad \text{(Claim 3)}\\
&= 2\cdot\sum_{j\geq i}\sum_{\substack{P\in\mathcal{P}_i\\ 2^{j-1}\leq \mathrm{sm}(P)}} \exp\left(-\frac{8\log(n)2^{j-1}}{2^i}\right) \qquad \text{(definition of } \rho)\\
&= 2\cdot\sum_{j\geq i} n^{-4\cdot 2^{j-i}}\cdot\left|\left\{\; P\in\mathcal{P}_i \;:\; 2^{j-1}\leq \mathrm{sm}(P)\;\right\}\right|\\
&\leq 2\cdot\sum_{j\geq i} n^{-4\cdot 2^{j-i}}\cdot n^{2\cdot 2^{j-i}} \qquad \text{(Theorem 5 with } \alpha = 2^{j-i})\\
&= 2\cdot\sum_{j\geq i} n^{-2\cdot 2^{j-i}} \;<\; 2\cdot\sum_{t\geq 2} n^{-t} \;<\; 3/n^2.
\end{aligned}$$

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 $\log(n)(3/n^2) < 1/n$ if n is sufficiently large. ■