# Advanced Algorithms 2012A
# Lecture 13 – Graph Compression and Cut Sparsifiers*

Robert Krauthgamer

The general idea is: given a graph on $n$ vertices, can we store the "relevant information" (think of distances or cuts) in a "compressed" form, i.e. using less than $O(|E(G)|)$ machine words? Often, the compression is achieved at the expense of some (actually, multiplicative) approximation of the relevant values. The stored information could take any form. A particular interesting representation is another graph $G'$ (say, smaller, i.e. with fewer edges), which would be useful if we later want to run some graph algorithms (on $G'$ instead of $G$). Sometimes we prefer a representation (data structure) that can answer queries about the relevant information quickly (say in $O(1)$ time).

Today we will focus on representing all cuts in a graph. For instance, if we want to solve some cut problem (say st-cut), the runtime may depend on number of edges, which may be as large as $n^2$. Can we approximate all the cuts in the input graph using a sparse graph (and compute this approximation quickly)?

## 1 Graph Sparsification for Cuts I

Let $G = (V, E)$ be an unweighted (multi-)graph, and suppose we want to approximate all cuts within factor $1 \pm \varepsilon$. (Since we allow parallel edges, we can actually handle "small" weights.) We shall write $G \leq G'$ if for every cut $(S, \bar{S})$, the capacity in $G$ at most that in $G'$.

**First try – subsampling.** Let's sample (i.e. keep) every edge independently with probability $p \in [0, 1]$. Denote the resulting graph $G' = (V, E')$. Consider a cut $(S, \bar{S})$, and suppose it's capacity (size) in $G$ is $c = |E(S, \bar{S})|$. Denote the capacity of the corresponding cut in $G'$ by a random variable $c' = |E'(S, \bar{S})|$. Then

$$\mathbb{E}[c'] = pc.$$

So in expectation, cuts are preserved up to scaling by a factor of $1/p$ (which can be "corrected" by giving every sampled edge capacity $1/p$). But is $c'$ likely to be close to its expectation? The answer is essentially yes, because it is the sum of *independent* indicators (one for each edge in the cut), as follows.

*These notes summarize the material covered in class, usually skipping proofs, details, examples and so forth, and possibly adding some remarks, or pointers. The exercises are for self-practice and need not be handed in. In the interest of brevity, most references and credits were omitted.

**Theorem [Chernoff-Hoeffding concentration bound].** Let $X = \sum_{i\in[n]} X_i$ where each $X_i \in [0,1]$ and $(X_i : i \in [n])$ are independently distributed. Then

1. $\forall t > 0, \quad \Pr[|X - \mathbb{E}X| > t] \le e^{-2t^2/n}$.
2. $\forall 0 < \varepsilon < 1, \quad \Pr[X < (1-\varepsilon)\mathbb{E}X] \le e^{-\varepsilon^2 \mathbb{E}X/2}$.
3. $\forall 0 < \varepsilon < 1, \quad \Pr[X > (1+\varepsilon)\mathbb{E}X] \le e^{-\varepsilon^2 \mathbb{E}X/3}$.
4. $\forall t > 2e\mathbb{E}X, \quad \Pr[X > t] \le 2^{-t}$.

Example/Motivation: $X$ is binomial $B(n, 1/2)$; then we can compare to Central Limit Theorem (tail of a Gaussian).

**Analysis of subsampling.** Using the concentration bound from above,

$$\Pr[c' > (1+\varepsilon)\mathbb{E}c'] \le e^{-\varepsilon^2 pc/3}.$$

Suppose we make sure that $p \ge \frac{3d \log n}{\varepsilon^2 c}$ for some fixed $d$ (say $d = 5$); then the RHS is $\le 1/n^d$. And since a similar bound applies to deviation in the other direction, we get

$$\Pr[c' \notin (1 \pm \varepsilon)\mathbb{E}c'] \le 2/n^d.$$

But is it possible to guarantee this approximation to *all* (exponentially many) cuts? The answer is yes, because in the number of small cuts is not too large, as shown in Theorem 1 below. We can then apply a union bound in "groups", each group having a number of events (cuts) that is inversely proportional to their probabilities.

**Theorem 1 [Karger].** Let $G$ be a graph on $n$ vertices, and let $\hat{c}$ denote its min-cut capacity. Then for every $\alpha \ge 1$, the number of cuts of capacity $\le \alpha c$ is at most $n^{2\alpha}$.

(Without proof.)

**Theorem 2 [Karger].** Let $G$ be a graph on $n$ vertices and min-cut capacity $\hat{c}$. Build $G'$ by including every edge from $G$ with probability $1 \le p \ge 12(d+1)\ln n/\varepsilon^2 \hat{c}$. Then with probability at least $1 - O(1/n^d)$, every cut in $G'$ has capacity within $1 \pm \varepsilon$ factor from its expectation.

Proof: shown in class.

Exer: Where did we use the fact that $G$ is unweighted? Extend the theorem to weighted graphs.

Later today we will need the following variant of Theorem 2.

**Theorem 3 [Karger].** Let $G = (V, E)$ be a graph and $X_e$ be independent random variables for $e \in E$ such that $X_e \in [0, M]$. Let $G(X_e)$ be the random graph obtained from $G$ by placing edge weights equal to $X_e$, and denote by $\hat{c}$ the minimum expected capacity over all cuts in $G(X_e)$. Then with probability $\geq 1 - O(1/n^d)$, we have every cut in $G(X_e)$ has capacity within $1 \pm \varepsilon$ factor of its expectation, where $\varepsilon = \sqrt{2(d+2)(M/\hat{c}) \ln n}$.

Example: if we are given a desired accuracy $\tilde{\varepsilon}$, we can set $p = \frac{2(d+2)\ln n}{\hat{c}\tilde{\varepsilon}^2}$. and let $X_e = 1/p$ with probability $p$, and $x_e = 0$ (i.e. non-edge) otherwise. Then $M = 1/p$ and we indeed get $\varepsilon = \sqrt{2(d+2)(M/\hat{c}) \ln n} = \tilde{\varepsilon}$. Notice this graph is exactly the one from previous theorem scaled by $1/p$ factor.

Exer: Prove this theorem (similarly to the previous one).

## 2 Graph Sparsification for Cuts II

The downside of the above result is that the number of edges decreases roughly by a factor of $\hat{c}/\ln n$, and in some cases it might still be quite dense (e.g. two cliques connected by a single edge, hence $\hat{c} = 1$). We now aim to overcome this.

**Theorem 4 [Benczur-Karger].** For every weighted graph $G = (V, E)$ on $n$ vertices and error parameter $\varepsilon > 0$, there is a weighted subgraph $G' = (V, E')$ with $O(\varepsilon^{-2} n \log n)$ edges such that $G' \in (1 \pm \varepsilon)G$. Moreover, $G'$ can be constructed in $O(|E| \log^2 n)$ time.

Such a graph $G'$ is called a $(1 + \varepsilon)$-cut sparsifier.

In class we proved a slightly weaker version, for unweighted graphs, with another $\log^2 n$ factor, and without the near-linear time algoritm.

**Main idea.** Sample edges non-uniformly, each edge $e$ with probability $p_e$ that is inversely proportional to its "connectivity" $c_e$. So "dense" regions will be sampled with smaller probability, thereby reducing the number of edges there more aggresively.

**Definitions of Connectivity.** A graph is *$k$-connected* if every cut in it has capacity $\geq k$. A *$k$-strong component* is a maximal vertex-induced subgraph that is $k$-connected.

It follows that the $k$-strong components partition the vertices of the graph, and that a $(k+1)$-strong components is a refinement of that partition.

The *strong connectivity of an edge* $e \in E$, denoted $c_e$, is the maximum value $k$ such that it is contained in a $k$-strong component. An edge is called *$k$-strong* if its strong connectivity is at least $k$, otherwise $k$-weak.

Note that strong connectivity differs from the usual definition of connectivity.

**Construction of sparsifier $G'$.** Let $q_\varepsilon = 4(d+2)\varepsilon^{-2}\ln n$. Then sample every edge $e \in E$ with probability $p_e = \min\{q_\varepsilon/c_e, 1\}$, in which case it is given weight $1/p_e$.

**Lemma 5.** If every edge $e$ is sampled with probability $\min\{q/c_e, 1\}$ then with at least $1-O(1/n^d)$ the resulting graph $G'$ has $O(qn)$ edges.

**Claim 5a.** A graph with total edge weight $\geq k(n-1)$ has a $k$-strong component (which may be the graph itself).

Exer: Prove the claim. (Hint: repeatedly find a cut of capacity $< k$.)

**Proof of Lemma 5.** We shall only consider the expected number of edges; the high-probability bound follows by Chernoff bound. The expected number of edges is at most $q\sum_{e\in E}(1/c_e)$.

To analyze this, consider a graph $\tilde{G}$, which is like $G$ but with edge weights $1/c_e$. It suffices to prove that this graph has total weight at most $n$. So let's assume to the contrary; then by the claim, it has a $\frac{n}{n-1}$-strong component $\tilde{F}$. Let $F$ be the corresponding subgraph of $G$, and let $\tilde{e}$ have minimum $c_{\tilde{e}}$ over $e \in F$.

By definition of $c_{\tilde{e}}$, $F$ cannot be more than $c_{\tilde{e}}$-connected, hence there is a cut $C$ of $F$ of capacity $\text{cap}(C, F) \leq c_{\tilde{e}}$. Consider the same cut $C$ in $\tilde{F}$; each edge of $C$ has capacity $1/c_e \leq 1/c_{\tilde{e}}$ (compared with 1 in $F$), thus

$$\text{cap}(C, \tilde{F}) \leq (1/c_{\tilde{e}})\,\text{cap}(C, F) \leq 1,$$

which contradicts the fact that $\tilde{F}$ is $\frac{n}{n-1}$-strong.

Exer: Complete the high-probability proof using Chernoff bound.

**Lemma 6** With high probability $G' \in (1 \pm \varepsilon \log|E|)G$.

Note that this indeed proves a weaker version of Theorem 4, by simply using Lemmas 5 and 6 with a smaller value $\varepsilon' = \varepsilon/\log|E|$.

The main idea is to use the uniform sampling (Theorem 3) in "parts". We thus divide the sampling process into phases: at phase $i = 0, 1, \ldots$ we flip the coins only for edges $e$ with $2^i \leq c_e < 2^{i+1}$.

**Proof** Decompose $G$ into graphs $G_i$ for $i = 0, 1, \ldots$, where $e \in G_i$ if $2^i \leq c_e < 2^{i+1}$. In the analysis (in order to use Theorem 3), at phase $i$ we actually consider the graph $G_{\geq i} = \cup_{j\geq i} G_j$; notice that it contains all $2^i$-strong components. At phase $i$, we sample edges of $G_{\geq i}$ as follows:

$$X_e^{(i)} = \begin{cases} 1/p_e \text{ w.p. } p_e, \text{ and } 0 \text{ otherwise} & \text{if } e \in G_i; \\ 1 & \text{otherwise (i.e. } e \in G_{\geq i+1}) \end{cases}$$

Recall that $p_e = \min\{q_\varepsilon/c_e, 1\}$. (Edges of levels lower than $i$ are not touched or considered at all.)

Consider a $2^i$-strong component $H$. Edges in $H \cap G_i$ are sampled with probability $p_e$, and edges in $H - G_i$ are always kept. Now in $H(X_e^{(i)})$, we can apply Theorem 3 with $\hat{c} \geq 2^i$ and $M = 2^{i+1}/q_\varepsilon$. Since $M/\hat{c} \leq 2q_\varepsilon$, we get that with high probability

$$H(X_e^{(i)}) \in (1 \pm \sqrt{2(d+2)(2/q_\varepsilon)\ln n})H = (1 \pm \varepsilon)H.$$

By combining this argument to the disjoint $2^i$-strong components in $G_{\geq i}$, we have $G_{\geq i}(X_e^{(i)}) \in (1 \pm \varepsilon)G_{\geq i}$.

Finally, we consider the entire graph $G$, incurring an error of $\varepsilon$ at each level $i$ (notice all $1 \leq c_e \leq |E|$):

$$\begin{aligned} G' &= \sum_{i=0}^{\log|E|} G_i(X_e^{(i)}) = \sum_i \left(G_{\geq i}(X_e^{(i)}) - G_{\geq i+1}\right) \\ &\in \sum_i (1 \pm \varepsilon) G_{\geq i} - \sum_i G_{\geq i+1} = \sum_i (1 \pm \varepsilon) G_i \pm \varepsilon \sum_i G_{\geq i+1} \\ &\in (1 \pm \varepsilon \log|E|)G. \end{aligned}$$

Theorem 7 now follows from Lemmas 5 and 6.

Exer: It is sometimes easier/faster to compute an approximation to $c_e$. So suppose we use in $p_e$ an approximation to $c_e$, say within factor 3, i.e., values $c'_e \in [c_e, 3c_e]$. Explain how how the theorem and analysis shown in class extend.