Sparsification

Fact (Karger)

$G$ has at most $cn^{2t/\lambda}$ cuts of size $t$ where $\lambda$ is the size of the min-cut and $c$ is some large constant.

Lemma

Let $G$ be an unweighted graph $G$ with minimum cut of size

$$\lambda > \lambda^* = 24\epsilon^{-2} \ln(2n^2 c).$$

Construct $G'$ by sampling each edge with probability $1/2$. Then,

$$\lambda_A(G') = (1 \pm \epsilon) \frac{\lambda_A(G)}{2} \quad \forall A \subset V$$

where $\lambda_A(\cdot)$ is the number of edges between $A$ and $V \setminus A$ in the graph.
Proof of Lemma

- Consider $A$ with $\lambda_A(G) = t$ and let $X = \lambda_A(G')$.
- Then $\mathbb{E}[X] = t/2$ and by an application of the Chernoff Bound,

$$P(|X - \mathbb{E}[X]| \geq \epsilon \mathbb{E}[X]) \leq 2 \exp(-\epsilon^2 t/6)$$

- Taking the union bound over all cuts gives,

$$\mathbb{P}[\lambda_A(G') \neq (1 \pm \epsilon)\lambda_A(G)/2 \text{ for some } A]$$

$$\leq \sum_{t \geq \lambda} \mathbb{P}[\lambda_A(G') \neq (1 \pm \epsilon)\lambda_A(G)/2 \text{ for some } A \text{ with } \lambda_A(G) = t]$$

$$\leq \sum_{t \geq \lambda} 2 \exp(-\epsilon^2 t/6) \cdot cn^{2t/\lambda}$$

$$= \sum_{t \geq \lambda} 2c \exp \left( \frac{2t \ln n}{\lambda} - \frac{\epsilon^2 t}{6} \right)$$

$$\leq \sum_{t \geq \lambda} 2c \exp \left( -\frac{\epsilon^2 t}{12} \right) \leq 2cn^2 \exp \left( -\frac{\epsilon^2 \lambda}{12} \right) \leq 1/n$$
Sparsification Algorithm

- Find “light” edges $L_0$ in $G$ where a set of edges is light if it’s removal leaves components with min-cut $\geq \lambda^*$. Let $G_1$ be formed by removing $L_0$ and sampling each remaining edge with probability $1/2$.

$$\lambda_A(G) = (1+\epsilon) 2\lambda_A(G_1) + \lambda_A(L_0)$$

- Find light edges $L_1$ in $G_1$. Let $G_2$ be formed by removing $L_1$ and sampling each remaining edge with probability $1/2$.

$$\lambda_A(G_1) = (1+\epsilon) 2\lambda_A(G_2) + \lambda_A(L_1)$$

and so

$$\lambda_A(G) = (1+\epsilon)^2 4\lambda_A(G_2) + 2\lambda_A(L_1) + \lambda_A(L_0)$$

- Next iteration,

$$\lambda_A(G) = (1+\epsilon)^3 8\lambda_A(G_3) + 4\lambda_A(L_2) + 2\lambda_A(L_1) + \lambda_A(L_0)$$

- Repeat $t = 2 \log n$ times: With high probability $G_t = \emptyset$ and so

$$\lambda_A(G) = (1+\epsilon)^t 2^t \lambda_A(L_t) + \ldots + 2\lambda_A(L_1) + \lambda_A(L_0)$$
We designed a sketch $A$ such that for any graph $G$, we can find a spanning forest $F$ from $A(G)$ with high probability.  

Construct $k$ independent spanning sketches $A_1(G), \ldots, A_k(G)$:

- $A_1(G)$ gives a spanning forest $F_1$ of $G$.
- $A_2(G) - A_2(F_1) = A_2(G - F_1)$ gives a spanning forest $F_2$ of $G - F_1$.
- $A_3(G) - A_3(F_1) - A_3(F_2) = A_3(G - F_1 - F_2)$ gives a spanning forest $F_3$ of $G - F_1 - F_2$.

Continue until we’ve found spanning forests $F_1, \ldots, F_k$.

Note that $F_1 \cup \ldots \cup F_k$ is $k$-connected iff $G$ is $k$-connected.

Furthermore, an edge $e$ is in a cut of size $\leq k - 1$ in $F_1 \cup \ldots \cup F_k$ iff it is in a cut of size $\leq k - 1$ in $G$.

Let’s call the overall sketch $B$. 

$k$-Edge Connectivity via Sketches
Finding Light Edges via $k$ connectivity sketch

- Define sets of edges $E_1, E_2, \ldots$ where

  \[ E_1 = \text{all edges in } G \text{ in a cut of size at most } \lambda^* - 1 \]

  \[ E_i = \text{all edges in } G - E_1 - E_2 - \ldots - E_i \text{ in a cut of size at most } \lambda^* - 1 \]

When the process terminates, $L = E_1 + E_2 + \ldots$ is set of light edges.

- We can find $E_1, E_2, \ldots$ from a $\lambda^*$ edge connectivity sketch $B(G)$:
  - $B(G)$ gives you $E_1$
  - $B(G) - B(E_1) = B(G - E_1)$ gives you $E_2$.
  - $B(G) - B(E_1) - B(E_2) = B(G - E_1 - E_2)$ gives you $E_3$ etc.
  - Continue until you’ve found $L$. 
Putting it all together

- Let $S_i$ be a sketch that samples each edge with probability $1/2^i$ where an edge is sampled using $S_i$ only if it is sampled using $S_{i-1}$.
- Sketch the data:
  \[ BS_0(G), BS_1(G), \ldots, BS_{2\log n}(G) \]
- Post-processing:
  1. $BS_0(G)$ gives $L_0$
  2. $BS_1(G)$ gives $L_1$ (ignore any edges already in $L_0$)
  3. $BS_2(G)$ gives $L_2$ (ignore any edges already in $L_0, L_1$)
  4. \ldots \ldots gives $L_t$ for $t = 2\log n$
- Return
  \[ L_0 + 2L_1 + 4L_2 + \ldots + 2^t L_t \]
- This is a $(1 + \epsilon)^{2\log n}$ sparsifier and the size of the sketches is $O(\epsilon^{-2} n \text{ polylog } n)$. Setting $\epsilon = \frac{\gamma}{2\log n}$ gives a $1 + \gamma$ sparsifier.