Cut Sparsifier of Arbitrary Graphs

- Let's check if the same subsampling also works for arbitrary graph $G = (V, E)$
- Same as before,

$$\mathbb{E}\left[|E(H)|\right] = p\,|E(G)|,$$

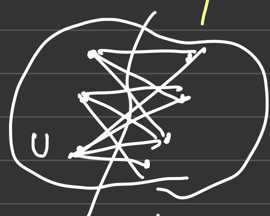
$$\forall U \subseteq V,\ \mathbb{E}\left[|E_H(U,\bar{U})|\right] = \frac{1}{p}|E_G(U,\bar{U})|$$

$\Rightarrow$ All cuts are preserved up to scaling by a factor $1/p$ in expectation !!

- Next, we use the concentration bound to analyze the followings :

$\Rightarrow \Pr\left\{ C_H > (1+\varepsilon)\,\mathbb{E}[C_H] \right\} \leq e^{-\frac{\varepsilon^2 \mathbb{E}[C_H]}{2}} = e^{-\frac{\varepsilon^2 p c_G}{2}}$

  - Note that this failure probability is exponentially decreaseing in c_G.
  - In the case of G being a complete graph, we can simply obtain the bound $\boxed{c_G \geq \frac{nq}{2}}$ ***



$G = K_n$

$c_G = |U| \cdot |\bar{U}|$ and we know that either $|U| = q \geq \frac{n}{2}$ or $|\bar{U}| = n - q \geq \frac{n}{2}$.

However, for the case of arbitrary graphs, we cannot obtain this kind of bound because no assumption can be made about the structure of G.

- Instead, we will use the fact about the number of cuts of small size.

Lemma 2 [Karger]: Let G be a graph with n vertices. Let c_G^* denote the size of minimum cuts of G. Then, for every $\alpha \geq 1$, the number of cuts of size at most αc_G^* is at most $n^{2\alpha}$.

Observation: A cut of size αc_G^* fails with probability at most $e^{-\frac{\varepsilon^2 p \alpha c_G^*}{3}}$

- Suppose among all cuts of size $\alpha \cdot c_G^*$ we want each cut to deviate from its mean more than a factor of $1+\varepsilon$, independently, with propability at most $\frac{1}{n^{\alpha d}}$. Then,

$$n^{2\alpha} \cdot e^{-\frac{\varepsilon^2 p \alpha c_G^*}{3}} \leq \frac{1}{n^{\alpha d}}$$

$$\Leftrightarrow \quad n^{2\alpha + \alpha d} \leq e^{\varepsilon^2 p \alpha c_G^*}$$

$$\Leftrightarrow \quad \cancel{\alpha}(d+2) \ln n \leq \varepsilon^2 p \cancel{\alpha} c_G^*$$

$$\Leftrightarrow \quad p \geq \frac{(d+2) \ln n}{\varepsilon^2 c_G^*}$$

Note that for each cut to deviate more than a factor of $(1 \pm \varepsilon)$, independenty, we have $\frac{2}{n^{\alpha d}}$ for the bouded probability.

- Now, with $p \geq \frac{(d+2)\ln n}{\varepsilon^2 c_G^*}$

$$\Pr\{\text{any cut fails}\} \leq \Pr\{\text{all cut fails independently}\}$$

$$\leq \int_1^{\infty} 2\left(n^{2\alpha} \cdot e^{-\frac{\varepsilon^2 p \cdot \alpha c_G^*}{3}}\right) d\alpha \leq \int_1^{\infty} \frac{2}{n^{d\alpha}} d\alpha = O\left(\frac{1}{n^d}\right)$$

Therefore, the subsampling succeeds w.h.p.

- In addition, given the value of c_G^* beforehand, the subsampling works in $O(m)$ time, and the resulting sparsifier H has

$$\mathbb{E}[|E(H)|] \geq \frac{(d+2)}{\varepsilon^2 c_G^*} \log n \cdot m$$

$$= \tilde{\Theta}\left(\frac{m}{\varepsilon^2 c_G^*}\right)$$

So, the number of edges decreases roughly by a factor of $\Theta\left(\frac{\log n}{c_G^*}\right)$ ← We shall discuss the limitation due to this result later.

Theorem 3:

The subsampling can construct a spasifier H of an arbitrary graph G, where $|E(H)| = \tilde{\Theta}\left(\frac{m}{\varepsilon^2 c_G^*}\right)$, with high probability of success.

Appendix : $\displaystyle\int_1^\infty \frac{1}{n^\alpha}\,d\alpha = \int_1^\infty n^{-\alpha}\,d\alpha$

$$= \left.\frac{n^{-\alpha}}{\ln n} + c\,\right|_1^\infty$$

$$= O\left(\frac{n^{-1}}{\ln n}\right) = O\left(\frac{1}{n\log n}\right)$$

$$= O\left(\frac{1}{n}\right)$$