CS 7880: Algorithms for Big Data (Fall'22)	**Northeastern University**

Lecture 20

November 22, 2022

Instructor: Soheil Behnezhad — *Scribe: Amir Azarmehr*

Disclaimer: *These notes have not been edited by the instructor.*

1 Edge-Degree Constrained Subgraphs

Graph sparsification is a common tool for dealing with big graphs. The idea is to come up with a sparse graph/subgraph that preserves certain properties of the original graph that we care about. The *edge-degree constrained subgraph (EDCS)*, introduced first by Bernstein and Stein [BS15], is such a sparsifier that robustly, in many senses, preserves the maximum matching of the graph approximately. In this lecture, we will cover its definition and some of its properties. The proofs we follow are due to a paper of Assadi and Bernstein [AB19].

1.1 Preliminaries

Theorem 1 (Extended Hall's Theorem)**.** *Let $G(L, R)$ be a bipartite graph. It holds that*

$$\mu(G) = \min_{A \subseteq L} |L \setminus A| + |N(A)|$$

Lemma 2 (Lovasz Local Lemma)**.** *Suppose $\mathcal{E}_1, \ldots, \mathcal{E}_t$ are t events such that $\Pr(\mathcal{E}_i) \leq p$ for all i, and $\mathcal{E}_i$ is mutually independent from all but k of the other events. If $p(k+1) < \frac{1}{e}$, then $\Pr\left(\cap_i \overline{\mathcal{E}_i}\right) > 0$.*

Definition 3. Given a graph G, a subgraph $H \subseteq G$ is an EDCS(G, β, β^-), where $\beta \geq \beta^- \geq 0$, if it has the following two properties:

(P1) $\forall (u, v) \in E_H$, we have $d_H(u) + d_H(v) \leq \beta$,

(P2) and $\forall (u, v) \in E_G \setminus E_H$, we have $d_H(u) + d_H(v) \geq (1 - \lambda)\beta$.

> **Remark.** A maximal matching is an EDCS with $\beta = 2$ and $\beta^- = 1$, which approximates maximum matching by a factor of $\frac{1}{2}$. As we will show, by letting β be a large enough constant and letting β^- be close enough to β, one can achieve a factor of $\frac{2}{3}$.

Lemma 4. *For any graph G and integers β and β^- such that $\beta^- \leq \beta - 1$, there exists a subgraph $H \subseteq G$ that is an* EDCS(G, β, β^-).

Proof. Consider the following process: Start with the subgraph H being empty. While H is not an EDCS(G, β, β^-), find an edge $e \in E_H$ that doesn't satisfy (P1) and remove it from H, or find an edge $e \in E_G \setminus E_H$ that doesn't satisfy (P2) and add it to H. We will show that this process will stop after at most $\mathcal{O}(n\beta^2)$ steps, meaning it would produce an EDCS.

To do so, we introduce a potential function:

$$\Phi(H) = \Phi_1(H) - \Phi_2(H),$$

where

$$\Phi_1(H) = \left(\beta - \frac{1}{2}\right) \sum_u d_H(u),$$

and

$$\Phi_2(H) = \sum_{(u,v)\in E_H} d_H(u) + d_H(v).$$

When $H = \emptyset$, the potential function is zero. At any point it holds that

$$\Phi(H) \leq \Phi_1(H) \leq n\beta\left(\beta - \frac{1}{2}\right)$$

To see this, note that in the beginning all vertices have degree zero in H. And when an edge is being added, both it endpoints must have degree smaller than $\beta^- \leq \beta - 1$. So the maximum degree is always bounded by β.

Now we will show that whenever we add or remove an edge, Φ increases by at least 1, hence the process will stop after at most $n\beta\left(\beta - \frac{1}{2}\right) = \mathcal{O}(n\beta^2)$ steps. First, let's say we are removing an edge (u, v), that violates (P1), from H. The sum of degrees will decrease by 2 and Φ_1 will decrease by $2\beta - 1$. Also, (u, v) is removed from the sum in Φ_2, and for every edge adjacent to (u, v), the edge degree will decrease by 1. So Φ_2 is decreased by $(d_H(u) + d_H(v)) + (d_H(u) - 1) + (d_H(v) - 1) \geq 2\beta$, and Φ will overall increase by at least 1.

Finally, consider adding an edge (u, v), that violates (P2), to H. The sum of degrees will increase by 2 and Φ_1 will increase by $2\beta - 1$. Similar to the previous case, (u, v) is added to the sum in Φ_2, and for every adjacent edge, the edge degree will increase by 1. So Φ_2 is increased by $(d_H(u)+d_H(v)+2)+d_H(u)+d_H(v) \leq 2\beta^- \leq 2\beta - 2$, and Φ will overall increase by at least 1. □

Lemma 5. *Let $H(L, R)$ be a bipartite graph such that for every edge $(u, v) \in E_H$ it holds that*

$$d_H(u) + d_H(v) \leq \beta.$$

Let $A \subseteq L$ be a vertex set of average degree $\bar{d}$, and let E be the set of outgoing edges of A. Then the average degree of $N(A)$ from the edges in E is at most $\beta - \bar{d}$.

Proof. Because of the upperbound on the edge degrees in H, we have:

$$\sum_{(u,v)\in E_H} d_H(u) + d_H(v) \leq \beta \cdot |E|$$

On the other hand we have ($d_E(u)$ denotes degree of u from the edges in E):

$$\begin{aligned}
\sum_{(u,v)\in E_H} d_H(u) + d_H(v) &= \sum_{u\in A} (d_H(u))^2 + \sum_{v\in N(A)} d_H(v) \cdot d_E(v) \\
&\geq \sum_{u\in A} (d_H(u))^2 + \sum_{v\in N(A)} (d_E(v))^2 \\
&\geq |A| \left(\frac{\sum_{u\in A} d_H(u)}{|A|}\right)^2 + |N(A)| \left(\frac{\sum_{v\in N(A)} d_E(v)}{|N(A)|}\right)^2 \qquad (*) \\
&= |A| \left(\frac{|E|}{|A|}\right)^2 + |N(A)| \left(\frac{|E|}{|N(A)|}\right)^2 \\
&= |E| \left(\frac{|E|}{|A|} + \frac{|E|}{|N(A)|}\right) \\
&= |E| \left(\bar{d} + \frac{|E|}{|N(A)|}\right)
\end{aligned}$$

Where $(*)$ follows from the fact that for variables with a fixed sum, the sum of squares is minimized when all the variables are equal. Putting this together with the inequality in the beginning, we conclude the proof:

$$\bar{d} + \frac{|E|}{|N(A)|} \leq \beta$$ □

1.2 Main Results

In the remaining section we will prove one of the most important properties of EDCS, that EDCS approximates maximum matching for both bipartite and non-bipartite graphs.

Theorem 6. *Let $G(L,R)$ be bipartite graph and $\varepsilon < \frac{1}{2}$. Let $\lambda \leq \frac{\varepsilon}{4}$, $\beta > 2\lambda^{-1}$, and $\beta^- \geq (1-\lambda)\beta$. Then for any $H \subseteq G$ that is an* $\mathrm{EDCS}(G,\beta,\beta^-)$*, it holds that*

$$\mu(G) \leq \left(\frac{3}{2} + \varepsilon\right)\mu(H).$$

Proof. Let A be a Hall's witness for H, i.e. such that $|L \setminus A| + |N(A)| = \mu(H)$. We define $\bar{A} = L \setminus A$, $B = N(A)$, and $\bar{B} = R \setminus B$. Note that by definition there are no edges between A and $\bar{B}$. Take a maximum matching M between A and $\bar{B}$ in G, and let S be the vertex set of M. We will state a few facts and then put them together to complete the proof. For now, take the following claim to be true:

Claim 7. *$M \geq \mu(G) - \mu(H)$.*

The vertices in S are perfectly matched with edges in $E_G \setminus E_H$. Hence, by (P2), S has average degree larger than $\beta^-/2 = \frac{1-\lambda}{2}\beta$ in H. More formally:

$$\begin{aligned}
\sum_{u\in S} d_H(u) &= \sum_{u\in S\cap L} d_H(u) + \sum_{v\in S\cap R} d_H(v) \\
&= \sum_{(u,v)\in M} d_H(u) + d_H(v) \\
&\geq |M| \cdot \beta^- \\
&= |S| \cdot \frac{\beta^-}{2}
\end{aligned}$$

Let E be the outgoing edges of S in H. By applying Lemma 5 to S in H, it follows that the average degree of $N_H(S)$ from the edges in E, is at most $\beta - \frac{\beta^-}{2} = \frac{1+\lambda}{2}\beta$. Also, $N_H(S) \subseteq \bar{A} \cup B$, so $|N_H(S)| \leq |\bar{A}| + |B| = \mu(H)$.

We are now ready to prove the theorem. Before giving a concise proof, consider the following argument. We have proven that the average degree of $N_H(S)$ from the edges in E is not much larger than that of S, so $|N_H(S)|$ must be almost as large as $|S|$. Hence:

$$2(\mu(G) - \mu(H)) \leq |S| \lesssim |N_H(S)| \leq |\bar{A}| + |B| = \mu(H)$$

Which gives:

$$\mu(G) \lesssim \frac{3}{2}\mu(H)$$

More formally:

$$\begin{aligned}
2(\mu(G) - \mu(H)) \cdot \frac{1-\lambda}{2}\beta &\leq |S| \cdot \frac{1-\lambda}{2}\beta \\
&\leq |E| \\
&\leq |N_H(S)| \cdot \frac{1+\lambda}{2}\beta \\
&\leq \mu(H) \cdot \frac{1+\lambda}{2}\beta
\end{aligned}$$

Moving the terms around we have:

$$2\mu(G) \leq \left(2 + \frac{1+\lambda}{1-\lambda}\right)\mu(H) \leq (3 + 2\varepsilon)\mu(H),$$

and the theorem follows. □

Proof of Claim 7. Notice that any maximum matching of H, covers all of $\bar{A}$, and $|N_H(A)|$ many of the vertices in A. So, any maximum matching of G, even if it covers all of $\bar{A}$, has to still cover $|N_H(A)| + (\mu(G) - \mu(H))$ vertices from A. And at least $\mu(G) - \mu(H)$ of them must be matched to outside of $N_H(A)$, which gives $|M| \geq \mu(G) - \mu(H)$.

More formally, let M^* be a maximum matching in G, and let $G^* = H \cup M^*$. By the choice of A, when have

$$\mu(H) = |L \setminus A| + |N_H(A)|,$$

and by the Extended Hall's Theorem, we have

$$\mu(G^*) \leq |L \setminus A| + |N_{G^*}(A)|.$$

Therefore,

$$\mu(G) - \mu(H) = \mu(G^*) - \mu(H) \leq |N_{G^*}(A)| - |N_H(A)| \leq |N_{G^* \setminus H}(A)|.$$

Notice that $G^* \setminus H$ is exactly the edges of M^* that are missing from H, Hence every vertex in $N_{G^* \setminus H}(A)$ is matched by an edge of M^* to A, which concludes the proof. □

Now we move on to the non-bipartite case.

Theorem 8. *Let G be a (possibly non-bipartite) graph, and let $\varepsilon < \frac{1}{2}$. Let $\lambda \leq \frac{\varepsilon}{32}$, $\beta \geq 100\lambda^{-2}\log(1/\lambda)$, and $\beta^- \geq (1-\lambda)\beta$. Then for any $H \subseteq G$ that is an* $\text{EDCS}(G, \beta, \beta^-)$*, it holds that*

$$\mu(G) \leq \left(\frac{3}{2} + \varepsilon\right)\mu(H).$$

Proof. To prove the theorem, we will introduce a bipartization $\tilde{G} = G[L, R]$ of G, where L and R are a partition of the vertices and $G[L, R]$ denotes the subgraph of G that includes the edges between L and R. This bipartization will have two properties:

1. $\mu(\tilde{G}) = \mu(G)$,
2. and $\tilde{H} = H \cap \tilde{G}$ is an $\text{EDCS}(\tilde{G}, \tilde{\beta}, \tilde{\beta}^-)$, where $\tilde{\beta} = \frac{1+4\lambda}{2}\beta$ and $\tilde{\beta}^- = \frac{1-5\lambda}{2}\beta$.

If such a bipartization exists, the theorem can be proven as follows:

$$\mu(H) \geq \mu(\tilde{H}) \geq \left(\frac{3}{2} + \varepsilon\right)\mu(\tilde{G}) \geq \left(\frac{3}{2} + \varepsilon\right)\mu(G)$$

Where the second inequality follows from Theorem 6 (the bipartite case).

To prove the existence, we use a probabilistic method. Fix a maximum matching M, and build L and R as follows:

1. For any edge (u, v), put u in either L or R with probability $\frac{1}{2}$, and put v on the other side.
2. For any unmatched vertex u, put it in L or R with probability $\frac{1}{2}$.

Notice that in this manner, $\tilde{G} = G[L, R]$ always includes M. So the first property holds, i.e. $\mu(\tilde{G}) = \mu(G)$. All it remains to show is that $\tilde{H}$ is an EDCS of $\tilde{G}$ with non-zero probability. Intuitively, the degrees of the vertices in $\tilde{H}$ are about half of what they were in H. So the inequalities (P1) and (P2) would hold for about half the value of β and β^-, which we have called $\tilde{\beta}$ and $\tilde{\beta}^-$.

To prove $\tilde{H}$ is an $\text{EDCS}(\tilde{G}, \tilde{\beta}, \tilde{\beta}^-)$ with non-zero probability, consider any vertex u that is unmatched in M (the matched case is similar). Assume, by symmetry, that u is in L. If any neighbours of u are matched together, exactly one of them will appear in R. Otherwise, they would appear in R independently with probability $\frac{1}{2}$. So we have:

$$\mathbb{E}\left[d_{\tilde{H}}(u)\right] = \frac{1}{2} d_H(u)$$

Now by the Chernoff bound (using β as an upperbound for the number of neighbours that are not matched together) we have:

$$\Pr\left(\left|d_{\tilde{H}}(u) - \mathbb{E}\left[d_{\tilde{H}}(u)\right]\right| > \lambda\beta\right) \leq \exp\left(-\frac{\lambda^2\beta^2}{3\beta}\right) \overset{(*)}{\leq} \exp(-4\log\beta) \leq \beta^{-4}$$

Where $(*)$ holds because $\beta \geq 12\lambda^{-2}\log\beta$. To see this note that the inequality holds for $\beta = 100\lambda^{-2}\log(1/\lambda)$ and $\frac{\beta}{\log\beta}$ is an increasing function.

Let $\mathcal{E}_u$ be the event that $\left|d_{\tilde{H}}(u) - \mathbb{E}\left[d_{\tilde{H}}(u)\right]\right| > \lambda\beta$. Notice that $\mathcal{E}_u$, only depends on the place of only some of the vertices, i.e. $\{u\} \cup N_H(u)$ and vertices matched to $\{u\} \cup N_H(u)$. Hence $\mathcal{E}_u$ would be independent of $\mathcal{E}_v$ if their distance is larger than 3, so $\mathcal{E}_u$ is independent of all but at most β^3 other events. And it holds that $\beta^3 \cdot \beta^{-4} = \beta^{-1} < \frac{1}{e}$. Applying the Lovasz Local Lemma, we get:

$$\Pr\left(\bigcap_u \overline{\mathcal{E}_u}\right) > 0,$$

That is there exists a bipartization $\tilde{G} = G[L, R]$ such that none of the $\mathcal{E}_u$'s occur. It is only left to show that $\tilde{H} = H \cap \tilde{G}$ is an EDCS$(\tilde{G}, \tilde{\beta}, \tilde{\beta}^-)$. Take any $(u, v) \in E_{\tilde{H}}$, we have:

$$d_{\tilde{H}}(u) + d_{\tilde{H}}(v) \leq \frac{1}{2}(d_H(u) + d_H(v)) + \lambda\beta \leq \frac{1+4\lambda}{\beta},$$

and for any $(u, v) \in E_{\tilde{G}} \setminus E_{\tilde{H}}$ we have:

$$d_{\tilde{H}}(u) + d_{\tilde{H}}(v) \geq \frac{1}{2}(d_H(u) + d_H(v)) - \lambda\beta \geq \frac{1}{2}\beta^- - \lambda\beta \geq \frac{1-5\lambda}{2}\beta \qquad \square$$

References

[BS15] Aaron Bernstein and Cliff Stein. "Fully Dynamic Matching in Bipartite Graphs". In: *Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I*. Ed. by Magnús M. Halldórsson et al. Vol. 9134. Lecture Notes in Computer Science. Springer, 2015, pp. 167–179. DOI: 10.1007/978-3-662-47672-7_14. URL: https://doi.org/10.1007/978-3-662-47672-7%5C_14.

[AB19] Sepehr Assadi and Aaron Bernstein. "Towards a unified theory of sparsification for matching problems". In: *2nd Symposium on Simplicity in Algorithms*. Vol. 69. OASIcs OpenAccess Ser. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2019, Art. No. 11, 20.