CSC2429, MAT1304: Circuit Complexity November 1, 2016

Lecture 8: AC^0 Circuit Size of SUB(G)

Instructor: Benjamin Rossman

Recall from last time:

- G is a fixed graph (no isolated vertices). X ranges over subgraphs of $G^{\uparrow n}$. SUB(G) is the problem: given X, does it contain a G-subgraph (of the form $G^{(\alpha)}$ for $\alpha \in [n]^{V(G)}$)?
- A *threshold weighting* is a function $\theta : E(G) \to [0,2]$ satisfying $\Delta_\theta(G) = 0$ and $\Delta_\theta(H) \geq 0$ for all $H \subseteq G$ where
$$\Delta_\theta(H) := |V(H)| - \sum_{e \in E(H)} \theta(e).$$
We say θ is *strict* if $\Delta_\theta(H) > 0$ for all $\emptyset \subset H \subset G$.
- $\mathbf{X}_\theta$ is the random subgraph of $G^{\uparrow n}$ where, independently for each $\{v^{(i)}, w^{(j)}\} \in E(G^{\uparrow n})$,
$$\mathbb{P}[\,\{v^{(i)}, w^{(j)}\} \in E(\mathbf{X}_\theta)\,] = n^{-\theta(\{v,w\})}.$$
We have $\mathbb{E}[\,\mathrm{sub}_H(\mathbf{X}_\theta)\,] = n^{\Delta_\theta(H)}$. If θ is strict, then $\mathrm{sub}_G(\mathbf{X}_\theta)$ is asymptotically Poisson(1) and $\mathrm{sub}_H(\mathbf{X}_\theta)$ is highly concentration around its mean for all $\emptyset \subset H \subset G$.

Convention. In the context of lower bounds for SUB(G), the adjective "AC^0" refers to circuits of depth $o(\frac{\log n}{\log\log n})$. (Our upper bounds will only require constant depth: $O(|E(G)|)$ in terms of G.)

Over the next two lectures, we will present results of Li, Razborov and Rossman (2014) that

1. characterize the average-case AC^0 circuit size of SUB(G) on $\mathbf{X}_\theta$ in terms of a combinatorial parameter $\kappa_\theta(G)$ (we give an $n^{\kappa_\theta(G)-o(1)}$ lower bound and an $n^{2\kappa_\theta(G)+O(1)}$ upper bound),
2. show that $\max_\theta \kappa_\theta(G) = \Omega(k/\log k)$ where k is the *tree-width* of G (we will define tree-width in the next lecture).

1 Patterns

In order to state the definition of $\kappa_\theta(G)$, we introduce the notion of *patterns* for a graph G. (Patterns were briefly discussed in Lecture 1 under the name "binary union trees".) Informally, a pattern is blueprint for constructing G by taking pairwise unions of subgraphs, starting out from individual edges.

Definition 1 (Patterns)**.**

- A *pattern for* G is a rooted binary tree A whose nodes are labeled by subgraphs of G such that
  - each leaf of A is labeled by a single-edge subgraph of G,

- each non-leaf of A is labeled by the union of the graphs labeling its two children,
- the root of A is labeled by G.

- The set of patterns for G is denoted by $\text{Pattern}(G)$. The set $\bigcup_{H \subseteq G} \text{Pattern}(H)$ of patterns for subgraphs of G is denoted by $\text{Pattern}(\subseteq G)$. Letters A, B will henceforth always represent patterns in $\text{Pattern}(\subseteq G)$.
- For $A \in \text{Pattern}(\subseteq G)$, the graph labeling the root of A is denoted by $G_A = (V_A, E_A)$. (That is, $E_A \subset E(G)$ is the set of edges that label leaves of A.)
- For $A, B \in \text{Pattern}(\subseteq G)$, we write $B \preceq A$ and say that B is a *sub-pattern* of A if B is a rooted subtree of A consisting of all descendants of a given node. Thus, the set $\{B : B \preceq A\}$ is in one-to-one correspondence with nodes of A; the set $\{G_B : B \preceq A\}$ is the set of subgraphs of G that label nodes of A.)
- For $A, B \in \text{Pattern}(\subseteq G)$, we denote by $\langle A, B \rangle \in \text{Pattern}(\subseteq G)$ the pattern consisting of a root with sub-patterns A and B as children. (Note that $G_{\langle A,B \rangle} = G_A \cup G_B$.)
- Patterns of size 1 (i.e. an isolated rooted labeled by a single edge of G) are said to be *atomic*.
- To streamline notation, for $A \in \text{Pattern}(\subseteq G)$, we write $\Delta_\theta(A)$ as a shorthand for $\Delta_\theta(G_A)$.

Example 2. We give two examples of patterns for path graphs. For integers $a < b$, let $P_{a,b}$ be the path of length $b - a$ with vertex set $\{a, a+1, \ldots, b\}$ and edge set $\{\{i, i+1\} : a \leq i < b\}$.

- The "recursive-doubling pattern" $\text{RD}_{a,b} \in \text{Pattern}(P_{a,b})$ is defined inductively as follows. If $b - a = 1$, then $\text{RD}_{a,b}$ is the atomic pattern labeled by the edge $\{a, b\}$. Otherwise,

$$\text{RD}_{a,b} := \langle \text{RD}_{a,c}, \text{RD}_{c,b} \rangle \text{ where } c = a + \lceil \tfrac{b-a}{2} \rceil.$$

- The "maximally-overlapping pattern" $\text{MO}_{a,b} \in \text{Pattern}(P_{a,b})$ is defined inductively as follows. If $b - a = 1$, then $\text{MO}_{a,b}$ is the atomic pattern labeled by the edge $\{a, b\}$. Otherwise,

$$\text{MO}_{a,b} := \langle \text{MO}_{a,b-1}, \text{MO}_{a+1,b} \rangle.$$

These two patterns will play an important role later on (when we prove AC^0 *formula* size lower bounds for $\text{SUB}(P_k)$).

2 The Parameter $\kappa_\theta(G)$

Definition 3. For a graph G and a threshold weighting θ, the parameter $\kappa_\theta(G)$ is defined by

$$\kappa_\theta(G) := \min_{A \in \text{Pattern}(G)} \max_{B \preceq A} \Delta_\theta(B).$$

In words, $\kappa_\theta(G)$ is the minimum over patterns A for G of the maximum over subgraphs $H \subseteq G$ that label nodes of A of the quantity $\Delta_\theta(H)$.

Note that $\kappa_\theta(G)$ is a constant between 0 and $|V(G)|$ (since $0 \le \Delta_\theta(H) \le |V(H)| \le |V(G)|$ for all $H \subseteq G$). Later on, we will see how to compute $\kappa_\theta(G)$ in the case $G = K_k$ and $\theta \equiv \frac{2}{k-1}$ (that is, where $\theta : E(G) \to [0,2]$ is the constant function $e \mapsto \frac{2}{k-1}$). (A priori, computing $\kappa_\theta(G)$ is no easy task: for $|E(G)| = k$, there roughly 2^{2^k} distinct patterns A and up to 2^k distinct graphs G_B for $B \preceq A$.)

Exercise 4. For any $k \ge 3$, in the case $G = C_k$ (the cycle of length k) and $\theta \equiv 1$, show that $\kappa_1(C_k) = 1$. [Hint: For $\emptyset \subseteq H \subseteq C_k$, what is $\Delta_1(H)$?]

Our next goal is the following theorem of Li, Razborov, Rossman (2014).

Theorem 5. *For every graph G and threshold weighting θ, the average-case AC^0 circuit size of $\mathrm{SUB}(G)$ on $\mathbf{X}_\theta$ is*

- **(lower bound)** *at least $n^{\kappa_\theta(G) - o(1)}$,*
- **(upper bound)** *at most $n^{2\kappa_\theta(G) + O(1)}$.*

After proving 5, we will show (in the next lecture)

Theorem 6. *For every graph G with tree-width $\ge k$,*

$$\Omega(k/\log k) \le \max_{\text{threshold weightings } \theta} \kappa_\theta(G) \le k+1.$$

As an immediate consequence of Theorems 5 and 6:

Corollary 7. *The AC^0 circuit size of $\mathrm{SUB}(G)$ is at least $n^{\Omega(\mathbf{tw}(G)/\log \mathbf{tw}(G))}$.*

Moreover, in special cases such as $G = K_k$ or when G is a constant-degree expander, we are able to show that $\max_\theta \kappa_\theta(G) = \Omega(|V(G)|)$, obtaining a nearly tight lower bound of $n^{\Omega(|V(G)|)}$ on the AC^0 circuit size of $\mathrm{SUB}(G)$.

3 Theorem 5: Upper Bound

We first present the $n^{2\kappa(G)+O(1)}$ upper bound of Theorem 5 by informally describing an algorithm for average-case $\mathrm{SUB}(G)$ on $\mathbf{X}_\theta$. A few additional ideas are required to implement this algorithm by AC^0 circuits of depth $O(|E(G)|)$; see (Li, Razborov, Rossman 2014) for details.

Informal description of the algorithm. Fix an arbitrary constant $\delta > 0$. We will say that $\mathbf{X}_\theta$ is "typical" if $\mathrm{sub}_H(\mathbf{X}_\theta) \le n^{\Delta_\theta(G)+\delta}$ for every $H \subseteq G$. By Markov's inequality,

$$\mathbb{P}[\ \mathbf{X}_\theta \text{ is not typical }] \le \sum_{H \subseteq G} \mathbb{P}\left[\ \mathrm{sub}_H(\mathbf{X}_\theta) > n^\delta\, \mathbb{E}[\ \mathrm{sub}_H(\mathbf{X}_\theta)\ ]\ \right] \le 2^{|E(G)|} \cdot n^{-\delta} \le o(1).$$

In other words, $\mathbf{X}_\theta$ is asymptotically almost surely typical (for any fixed constant $\delta > 0$).

We now describe an $n^{2\kappa_\theta(G)+O(1)}$ time algorithm which, given *any* typical $X \subseteq G^{\uparrow n}$, determines whether or not $\mathrm{sub}_G(X) \ge 1$. By definition of $\kappa_\theta(G)$, there exists a pattern $A \in \mathrm{Pattern}(G)$ with

$$\kappa_\theta(G) = \max_{B \preceq A} \Delta_\theta(B).$$

Our algorithm does the following: for each sub-pattern $B \preceq A$, we will compute the list L_B of all H-subgraphs in X. By assumption of typicality, each list L_B contains $\le n^{\Delta_\theta(G)+\delta}$ items. These lists are constructed by induction on sub-patterns $B \preceq A$:

- For each leaf $B \preceq A$ (labeled by an edge $\{v, w\} \in E(G)$), we compute L_B by checking all n^2 potential edge of the $\{v^{(i)}, w^{(j)}\}$.
- For each non-leaf $B = \langle B_1, B_2 \rangle \preceq A$ with $H_1 = G_{B_1}$ and $H_2 = G_{B_2}$, we compute L_B by checking, for each pairs $H_1^{(\alpha_1)} \in L_{B_1}$ and $H_2^{(\alpha_2)} \in L_{B_2}$, whether or not $H_1^{(\alpha_1)} \cup H_2^{(\alpha_2)}$ is a valid $(H_1 \cup H_2)$-subgraph of X (i.e. whether $\alpha_1(v) = \alpha_2(v)$ for all $v \in V(H_1) \cap V(H_2)$). This merging of lists L_{B_1} and L_{B_2} requires inspecting $|L_{B_1}| \cdot |L_{B_2}| \leq n^{\Delta_\theta(B_1)+\Delta_\theta(B_2)+2\delta} \leq n^{2\kappa_\theta(G)+2\delta}$ pairs.

The final list L_A lets us determine whether $\text{sub}_G(X) \geq 1$. The total complexity is at most $n^{2\kappa_\theta(G)+O(1)}$ (since $\delta > 0$ is a constant, it is absorbed in the $O(1)$).

Remark 8. This algorithm gives an upper bound of $n^{\kappa_\theta^{\text{upper}}(G)+O(1)}$ where

$$\kappa_\theta^{\text{upper}}(G) := \min_{A \in \text{Pattern}(G)} \max_{\langle B_1, B_2 \rangle \preceq A} \Big(\Delta_\theta(B_1) + \Delta_\theta(B_2) \Big).$$

(Clearly $\kappa_\theta(G) \leq \kappa_\theta^{\text{upper}}(G) \leq 2\kappa_\theta(G)$.) I suspect that $n^{\kappa_\theta(G)+O(1)}$ is the true upper bound. Perhaps this can be achieved by merging lists L_{B_1} and L_{B_2} in a more efficient fashion. *Let me know if you have any ideas!*

4 Dual Expression for $\kappa_\theta(G)$

Before turning attention to the lower bound of Theorem 5, we remark on dual for $\kappa_\theta(G)$.

Definition 9. A *hitting set* for G is a family $\mathcal{H} \subseteq \{\text{subgraphs of } G\}$ such that for every $A \in \text{Pattern}(G)$, there exists $B \preceq A$ with $G_B \in \mathcal{H}$. One example of a hitting set, for any graph G, is the family $\{H \subseteq G : \frac{1}{3}|V(G)| \leq |V(H)| \leq \frac{2}{3}|V(G)|\}$.

Lemma 10. $\kappa_\theta(G) = \max_{\text{hitting set } \mathcal{H}} \min_{H \in \mathcal{H}} \Delta_\theta(H)$.

Proof. First, we show that

$$\max_{\text{hitting set } \mathcal{H}} \min_{H \in \mathcal{H}} \Delta_\theta(H) \leq \min_{A \in \text{Pattern}(G)} \max_{B \preceq A} \Delta_\theta(B) \ (=: \kappa_\theta(G)).$$

Indeed, for *any* hitting set $\mathcal{H}$ and pattern A, we have

$$\min_{H \in \mathcal{H}} \Delta_\theta(H) \leq \max_{B \preceq A} \Delta_\theta(B).$$

This inequality is witnessed by any choice of $B \preceq A$ with $G_B \in \mathcal{H}$.

For the opposite inequality, consider the specific hitting set

$$\mathcal{H}_0 := \{G_B : B \preceq A \in \text{Pattern}(G) \text{ and } \Delta_\theta(B) = \max_{B' \preceq A} \Delta_\theta(B')\}.$$

This family $\mathcal{H}_0$ is clearly a hitting set for G. Fix any $A_0 \in \text{Pattern}(G)$ which minimizes $\max_{B \preceq A_0} \Delta(B)$. Next, fix any $B_0 \preceq A_0$ which maximizes $\Delta(B_0)$. We have $\kappa_\theta(G) = \Delta(B_0)$ (by definition of $\kappa_\theta(G)$). Since $G_{B_0} \in \mathcal{H}_0$ (by definition of $\mathcal{H}_0$), we have

$$\kappa_\theta(G) = \Delta_\theta(B_0) \leq \min_{H \in \mathcal{H}_0} \Delta_\theta(H) \leq \max_{\text{hitting set } \mathcal{H}} \min_{H \in \mathcal{H}} \Delta_\theta(H).$$

□

We next show how to bound $\kappa_\theta(G)$: each pattern A gives an upper bound on $\kappa_\theta(G)$, while each hitting set $\mathcal{H}$ gives a lower bound. In the case of $G = K_k$ and $\theta \equiv \frac{2}{k-1}$, we can show that two bounds meet at around $k/4$. (This calculation serves to illustrate the definition of $\kappa_\theta(G)$.)

Proposition 11. *$\kappa_\theta(K_k) = \dfrac{k}{4} + O(1)$ where $\theta \equiv \frac{2}{k-1}$*

Proof. For $j \in \{2, \dots, k\}$ and $\lambda := j/k$, we have

$$\Delta_\theta(K_j) = j - \frac{2}{k-1}\binom{j}{2} = j\left(1 - \frac{j-1}{k-1}\right) = \lambda(1-\lambda)k + O(1/k).$$

Upper bound. Let $e_1, \dots, e_{\binom{k}{2}}$ be the edges of K_k in lexicographic order: $\{1,2\}$, $\{1,3\}$, $\{2,3\}$, $\{1,4\}$, $\{2,4\}$, $\{3,4\}$, $\{1,5\}$, ... Let A be the pattern which adds these edge one-by-one: $A := A_{\binom{k}{2}}$ where $A_2 := \langle e_1, e_2\rangle$ and $A_i = \langle A_{i-1}, e_i\rangle$ for all $3 \le i \le \binom{k}{2}$. Observe that, for each $B \preceq A$, either G_B is a single edge (in which case $\Delta_\theta(B) = 2 - \frac{2}{k-1}$), or there exists $j \in \{2, \dots, k-1\}$ such that $K_j \subseteq G_B \subseteq K_{j+1}$. In the latter case, we have

$$\Delta_\theta(B) = |V_B| - \sum_{e \in E_B} \theta(e) = j + 1 - \binom{j}{2}\frac{2}{k-1}s = \Delta_\theta(K_j) + 1.$$

Therefore,

$$\kappa_\theta(K_k) \le \max_{B \preceq A} \Delta_\theta(B) \le \max\left\{2, \max_{2 \le j \le k-1} \Delta_\theta(K_j) + 1\right\} \le \max_{\lambda \in [0,1]} \lambda(1-\lambda)k + O(1) = \frac{k}{4} + O(1).$$

Lower bound (1st attempt). Consider the hitting set

$$\mathcal{H} = \left\{H \subseteq K_k : \frac{k}{3} \le |V(H)| \le \frac{2k}{3}\right\}.$$

Observe that $\Delta_\theta(H) \ge \Delta_\theta(K_{|V(H)|})$ for every $H \subseteq G$. Therefore (using the dual expression for $\kappa_\theta(K_k)$)

$$\kappa_\theta(K_k) \ge \max_{H \in \mathcal{H}} \Delta_\theta(H) \ge \min_{\lambda \in [\frac{1}{3}, \frac{2}{3}]} \lambda(1-\lambda)k = \frac{2k}{9}.$$

Lower bound (2nd attempt). Consider the hitting set

$$\mathcal{H} = \left\{H \subseteq K_k : |V(H)| > \frac{k}{2} \text{ and } H = H_1 \cup H_2 \text{ where } |V(H_1)|, |V(H_2)| \le \frac{k}{2}\right\}.$$

A straightforward calculation shows that $\Delta_\theta(H) \ge k/4$ for all $H \in \mathcal{H}$. (This is minimal when $H = K_{\lceil k/2 \rceil} - \{\text{single edge}\}$.) □

5 Key Technical Lemma: "H-Shaped Sensitivity" of AC^0 Functions

In order to prove the lower bound of Theorem 5, we require a lemma on Boolean functions

$$f : \{0,1\}^{E(G^{\uparrow n})} \ (\cong \{\text{subgraphs of } G^{\uparrow n}\}) \to \{0,1\}$$

that are computable by AC^0 circuits (specifically, circuits of depth $o(\frac{\log n}{\log\log n})$ and size $n^{O(\log\log n)}$).

Definition 12. For graphs $X, S \subseteq G^{\uparrow n}$, we define a restriction $R_{X,S} : E(G^{\uparrow n}) \to \{0,1,*\}$ by

$$R_{X,S} : e \mapsto \begin{cases} * & \text{if } e \in E(S), \\ 1 & \text{if } e \in E(X) \setminus E(S), \\ 0 & \text{otherwise.} \end{cases}$$

That is, S specifies the stars of the restriction (i.e. the unrestricted variables) and X (or rather $X \setminus S$) specifies the setting of restricted variables. We may apply the restriction $R_{X,S}$ to f in the usual way to obtain a Boolean function

$$f{\restriction}R_{X,S} : \{0,1\}^{E(S)} \ (\cong \{\text{subgraphs of } S\}) \to \{0,1\}.$$

We now fix a graph G and a strict threshold weighting θ, as well as nonempty proper subgraph $\emptyset \subset H \subset G$. We are interested in the random restriction $R_{\mathbf{X},\mathbf{H}}$ where

- $\mathbf{X} := \mathbf{X}_\theta$ (we drop the subscript θ to streamline notation in what follows),
- $\mathbf{H} \subseteq G^{\uparrow n}$ is a uniform random H-subgraph of $G^{\uparrow n}$ (independent of $\mathbf{X}$), that is, $\mathbf{H}$ has the same distribution as $H^{(\boldsymbol{\alpha})}$ for uniform random $\boldsymbol{\alpha} \in [n]^{V(H)}$.

Our main lemma states that if f is AC^0-computable, then the probability that $f{\restriction}R_{\mathbf{X},\mathbf{H}}$ depends on all of its variables (i.e. all coordinates in $E(\mathbf{H})$) is at most $1/n^{\Delta_\theta(H)-o(1)}$.

Convention. We stick to the habit of denoting random objects in boldface.

Lemma 13 (Key Technical Lemma). *Suppose $f : \{0,1\}^{E(G^{\uparrow n})} \to \{0,1\}$ is computed by circuits of depth $o(\frac{\log n}{\log\log n})$ and size $n^{O(\log\log n)}$. Then*

$$\mathbb{P}[\ \mathrm{ALL}(f{\restriction}R_{\mathbf{X},\mathbf{H}})\ ] \le n^{-\Delta_\theta(H)+o(1)}.$$

where $\mathrm{ALL}(f{\restriction}R_{\mathbf{X},\mathbf{H}})$ is the event that the restricted function $f{\restriction}R_{\mathbf{X},\mathbf{H}} : \{0,1\}^{E(\mathbf{H})} \to \{0,1\}$ depends on all of its variables.

For those who are familiar with the notion of average sensitivity, $\mathbb{P}[\ \mathrm{ALL}(f{\restriction}R_{\mathbf{X},\mathbf{H}})\ ]$ may be seen as the "average H-shaped sensitivity" of f under the the distribution $\mathbf{X}$. In the case where H is a single edge, this essentially coincides with the usual notion of average sensitivity (up to a normalization factor).

Proof. Fix any small constant $\delta > 0$ (independent of n). We prove the lemma by showing that

$$\mathbb{P}[\ \mathrm{ALL}(f{\restriction}R_{\mathbf{X},\mathbf{H}})\ ] \le n^{-\omega(\log\log n)} + O(n^{-\Delta_\theta(H)+\delta\cdot(|V(H)|+|E(H)|)}). \tag{1}$$

The first term is $n^{-\omega(\log\log n)}$ for any fixed choice of $\delta > 0$. Since $\delta > 0$ may be chosen arbitrary small, the second term may be replaced by $n^{-\Delta_\theta(H)+o(1)}$, which yields the lemma.

We sample $\mathbf{H}$ using a two-step process. First, (independent of $\mathbf{X}$) we generate an auxiliary random graph $\mathbf{S} \subseteq G^{\uparrow n}$ under the product distribution where where

$$\mathbb{P}[\, e \in E(\mathbf{S})\,] = n^{-\theta(\{v,w\})-\delta}$$

for each $\{v^{(i)}, w^{(j)}\} \in E(G^{\uparrow n})$. (Each edge probability in $\mathbf{S}$ is smaller than the corresponding edge probability in $\mathbf{X}$ by a factor of n^δ.) Having sampled $\mathbf{S}$, we then generate $\mathbf{H}$ as follows: in the (overwhelmingly likely) event that $\mathrm{sub}_H(\mathbf{S}) \geq 1$, let $\mathbf{H}$ be a uniform random H-subgraph of $\mathbf{S}$; otherwise, let $\mathbf{H}$ be a uniform random H-subgraph of $G^{\uparrow n}$. Note that $\mathbf{H}$ is independent of $\mathbf{X}$ (since $\mathbf{S}$ is independent of $\mathbf{X}$).

Observe that

$$\mathbb{E}[\, \mathrm{sub}_H(\mathbf{S})\,] = n^{|V(H)|-\sum_{e\in E(H)}(\theta(e)+\delta)} = n^{\Delta_\theta(H)-\delta|E(H)|}.$$

We may assume that the constant $\delta > 0$ is chosen small enough so that $\Delta_\theta(H') - \delta|E(H')| > 0$ for all $\emptyset \subset H' \subset H$. Under this assumption, the random variable $\mathrm{sub}_H(\mathbf{S})$ will be highly concentrated around its mean. In particular, we have the following claim (essentially Theorem 13 of the last lecture):

Claim 14. $\mathbb{P}[\, \mathrm{sub}_H(\mathbf{S}) < \frac{1}{2}n^{\Delta_\theta(H)-\delta|E(H)|}\,] \leq \exp(-n^{\Omega(1)})$.

(The constant $\Omega(1)$ in Claim 14 will depend on $\min_{\emptyset\subset H'\subset H} \Delta_\theta(H') - \delta|E(H')|$, which is why we require choosing small enough δ so that this is positive.) We omit the proof of Claim 14, which is similar to well-known concentration-of-measure inequalities in random graph theory.

Recall that $\mathcal{D}(\cdot)$ denotes decision-tree depth. Using Håstad's Switching Lemma, we will show:

Claim 15. $\mathbb{P}[\, \mathcal{D}(f\restriction R_{\mathbf{X},\mathbf{S}}) > \delta \log n\,] \leq n^{-\omega(\log\log n)}$.

In order not to interrupt the flow of the main argument, we defer the proof of Claim 15 until the end of the proof of Lemma 13.

The third and final claim we require to establish inequality (1) is:

Claim 16. *For all $X, S \subseteq G^{\uparrow n}$,*

$$|\{\alpha \in [n]^{V(H)} : H^{(\alpha)} \subseteq S \text{ and } \mathrm{ALL}(f\restriction R_{X,H^{(\alpha)}})\}| \leq 2^{\mathcal{D}(f\restriction R_{X,S})\cdot|V(H)|}. \tag{2}$$

In particular, if $\mathcal{D}(f\restriction R_{X,S}) \leq \delta\log n$ and $\mathrm{sub}_H(S) \geq \frac{1}{2}n^{-\Delta_\theta(H)+\delta|E(H)|}$, then

$$\mathop{\mathbb{P}}_{\boldsymbol{\alpha}\in[n]^{V(H)}}[\, \mathrm{ALL}(f\restriction R_{X,H^{(\boldsymbol{\alpha})}}) \mid H^{(\boldsymbol{\alpha})} \subseteq S\,] \leq 2n^{-\Delta_\theta(H)+\delta(|E(H)|+|V(H)|)}. \tag{3}$$

To see why (2) holds, consider any $X, S \subseteq G^{\uparrow n}$ and let $J \subseteq E(S)$ be the set of variables on which $f\restriction R_{X,S} : \{0,1\}^{E(S)} \to \{0,1\}$ depends. Note that $|J| \leq 2^{\mathcal{D}(f\restriction R_{X,S})}$ (since every depth k decision tree depends on $\leq 2^k$ variables). For each $v \in V(H)$, define $I_v \subseteq [n]$ by

$$I_v := \big\{i \in [n] : \{v^{(i)}, w^{(j)}\} \in J \text{ for some } w \in V(H) \text{ and } j \in [n]\big\}.$$

Now observe that, for all $\alpha \in [n]^{V(H)}$ with $H^{(\alpha)} \subseteq S$, we have

$$\begin{aligned}\mathrm{ALL}(f\restriction R_{X,H^{(\alpha)}}) &\iff f\restriction R_{X,H^{(\alpha)}} \text{ depends on all variables in } E(H^{(\alpha)})\\ &\implies E(H^{(\alpha)}) \subseteq J\\ &\implies \bigvee_{v\in V(H)} (\alpha_v \in I_v).\end{aligned}$$

It follows that

$$\begin{aligned}|\{\alpha \in [n]^{V(H)} : H^{(\alpha)} \subseteq S \text{ and } \mathrm{ALL}(f\restriction R_{X,H^{(\alpha)}})\}| &\le |\{\alpha \in [n]^{V(H)} : \textstyle\bigvee_{v\in V(H)} (\alpha_v \in I_v)\}|\\ &\le \prod_{v\in V(H)} |I_v|\\ &\le |J|^{|V(H)|} \quad (\text{since } |I_v| \le |J| \text{ for all } v \in V(H))\\ &\le 2^{\mathcal{D}(f\restriction R_{X,S})\cdot|V(H)|}.\end{aligned}$$

This proves inequality (2). Inequality (3) follows, since if we assume that $\mathcal{D}(f\restriction R_{X,S}) \le \delta \log n$ and $\mathrm{sub}_H(S) \ge \frac{1}{2}n^{-\Delta_\theta(H)+\delta|E(H)|}$, then

$$\begin{aligned}\mathop{\mathbb{P}}_{\boldsymbol{\alpha}\in[n]^{V(H)}}[\ \mathrm{ALL}(f\restriction R_{X,H^{(\boldsymbol{\alpha})}}) \mid H^{(\boldsymbol{\alpha})} \subseteq S\ ] &= \frac{|\{\alpha \in [n]^{V(H)} : H^{(\alpha)} \subseteq S \text{ and } \mathrm{ALL}(f\restriction R_{X,H^{(\alpha)}})\}|}{\mathrm{sub}_H(S)}\\ &\le \frac{2^{\mathcal{D}(f\restriction R_{X,S})\cdot|V(H)|}}{\mathrm{sub}_H(S)}\\ &\le \frac{2^{\delta\log n\cdot|V(H)|}}{\frac{1}{2}n^{\Delta_\theta(H)-\delta|E(H)|}}\\ &= 2n^{-\Delta_\theta(H)+\delta(|E(H)|+|V(H)|)}.\end{aligned}$$

Equipped with Claims 14, 15, 16, we finish the proof of Lemma 13 by deriving inequality (1):

$$\begin{aligned}\mathbb{P}[\ \mathrm{ALL}(f\restriction R_{\mathbf{X},\mathbf{H}})\ ] &\le \mathbb{P}[\ \mathrm{sub}_H(\mathbf{S}) < \tfrac{1}{2}n^{\Delta_\theta(H)-\delta|E(H)|}\ ]\\ &\quad + \mathbb{P}[\ \mathcal{D}(f\restriction R_{\mathbf{X},\mathbf{S}}) > \delta\log n\ ]\\ &\quad + \mathbb{P}[\ \mathrm{ALL}(f\restriction R_{\mathbf{X},\mathbf{H}}) \mid \mathrm{sub}_H(\mathbf{S}) \ge \tfrac{1}{2}n^{\Delta_\theta(H)-\delta|E(H)|} \ \&\ \mathcal{D}(f\restriction R_{\mathbf{X},\mathbf{S}}) \le \delta\log n\ ]\\ &\le \exp(-n^{\Omega(1)}) \qquad (\text{by Claim 14})\\ &\quad + n^{-\omega(\log\log n)} \qquad (\text{by Claim 15})\\ &\quad + 2n^{-\Delta_\theta(H)+\delta(|E(H)|+|V(H)|)} \qquad (\text{by Claim 16})\\ &= n^{-\omega(\log\log n)} + O(n^{-\Delta_\theta(H)+\delta(|E(H)|+|V(H)|)})\\ &= n^{-\Delta_\theta(H)+o(1)} \qquad (\text{taking } \delta > 0 \text{ arbitrarily small}).\end{aligned}$$

Proof of Claim 15. It remains to prove Claim 15 using Håstad's Switching Lemma. Let C be a circuit of size $s = n^{O(\log\log n)}$ and depth $d = o(\frac{\log n}{\log\log n})$ computing f. Note that the random restriction $R_{\mathbf{X},\mathbf{S}} : E(G^{\uparrow n}) \to \{0,1,*\}$ has a product distribution (i.e. values $R_{\mathbf{X},\mathbf{S}}(e)$ are independent)

where, for each $e = \{v^{(i)}, w^{(j)}\} \in E(G^{\uparrow n})$, letting $p_e := n^{-\theta(\{v,w\})}$, we have

$$\begin{aligned}
\mathbb{P}[\ R_{\mathbf{X},\mathbf{S}}(e) = * \ ] &= p_e n^{-\delta}, \\
\mathbb{P}[\ R_{\mathbf{X},\mathbf{S}}(e) = 1 \ ] &= (1 - p_e n^{-\delta}) p_e \\
\mathbb{P}[\ R_{\mathbf{X},\mathbf{S}}(e) = 0 \ ] &= 1 - (1 - p_e n^{-\delta}) p_e.
\end{aligned}$$

In order to use the Switching Lemma, we generate $R_{\mathbf{X},\mathbf{S}}$ via sequence of $d+1$ auxiliary random restrictions $\boldsymbol{\varrho}_0, \ldots, \boldsymbol{\varrho}_d$ defined as follows:

- Let $\boldsymbol{\varrho}_0 : E(G^{\uparrow n}) \to \{0, 1, *\}$ be the random restriction where, independently for each $e \in E(G^{\uparrow n})$,

  $$\begin{aligned}
  \mathbb{P}[\ \boldsymbol{\varrho}_0(e) = * \ ] &= p_e, \\
  \mathbb{P}[\ \boldsymbol{\varrho}_0(e) = 1 \ ] &= q_e \qquad \text{(to be determined, but roughly } \tfrac{1}{2} p_e\text{)}, \\
  \mathbb{P}[\ \boldsymbol{\varrho}_0(e) = 0 \ ] &= 1 - p_e - q_e.
  \end{aligned}$$

  (Nota bene: We shall not apply the Switching Lemma with respect to $\boldsymbol{\varrho}_0$. This initial restriction simply allows us to generate $R_{\mathbf{X},\mathbf{S}}$ as the eventual $\boldsymbol{\varrho}_d$.)

- Inductively, for $i = 1, \ldots, d$, let $\boldsymbol{\varrho}_i : E(G^{\uparrow n}) \to \{0, 1, *\}$ be the random restriction where, independently for each $e \in E(G^{\uparrow n})$,

  $$\begin{aligned}
  \mathbb{P}[\ \boldsymbol{\varrho}_i(e) = \boldsymbol{\varrho}_{i-1}(e) \mid \boldsymbol{\varrho}_{i-1}(e) \neq * \ ] &= 1, \\
  \mathbb{P}[\ \boldsymbol{\varrho}_i(e) = * \mid \boldsymbol{\varrho}_{i-1}(e) = * \ ] &= n^{-\delta/d}, \\
  \mathbb{P}[\ \boldsymbol{\varrho}_i(e) = 1 \mid \boldsymbol{\varrho}_{i-1}(e) = * \ ] &= \tfrac{1}{2}(1 - n^{-\delta/d}), \\
  \mathbb{P}[\ \boldsymbol{\varrho}_i(e) = 0 \mid \boldsymbol{\varrho}_{i-1}(e) = * \ ] &= \tfrac{1}{2}(1 - n^{-\delta/d}).
  \end{aligned}$$

  In other words, if we condition on $\boldsymbol{\varrho}_{i-1} = \varrho_{i-1}$ for any given restriction ϱ_{i-1}, then $\boldsymbol{\varrho}_i$ is a random extension of ϱ_{i-1} which sets each $e \in \varrho_{i-1}^{-1}(*)$ independently to $*$ with probability $n^{-\delta/d}$ and to 0 or 1 with probability $\frac{1}{2}(1 - n^{-\delta/d})$ each. In particular, conditioned on $\boldsymbol{\varrho}_{i-1} = \varrho_{i-1}$, the random restriction $\boldsymbol{\varrho}_i$ is *unbiased* with respect to 0's and 1's. By Håstad's Switching Lemma, it follow that if g is the AND or OR of Boolean functions $h_1, \ldots, h_m : \{0,1\}^{E(G^{\uparrow n})} \to \{0, 1, *\}$, then for all $k, \ell \in \mathbb{N}$,

  $$\mathbb{P}\left[\mathcal{D}(g{\restriction}\boldsymbol{\varrho}_i) \geq \ell \;\middle|\; \bigwedge_{j \in 1}^{m} \mathcal{D}(h_j{\restriction}\boldsymbol{\varrho}_{i-1}) \leq k \right] \leq (5 n^{-\delta/d} k)^{\ell}. \tag{4}$$

Let's check that $\boldsymbol{\varrho}_d$ has the same distribution as $R_{\mathbf{X},\mathbf{S}}$ (for an appropriate choice of q_e). First, we check

$$\begin{aligned}
\mathbb{P}[\ \boldsymbol{\varrho}_d(e) = * \ ] &= \mathbb{P}[\ \bigwedge_{i=0}^{d} (\boldsymbol{\varrho}_i(e) = *) \ ] \\
&= \mathbb{P}[\ \boldsymbol{\varrho}_0(e) = * \ ] \prod_{i=1}^{d} \mathbb{P}[\ \boldsymbol{\varrho}_i(e) = * \mid \boldsymbol{\varrho}_{i-1}(e) = * \ ] \\
&= p_e (n^{-\delta/d})^d \\
&= p_e n^{-\delta} \\
&= \mathbb{P}[\ R_{\mathbf{X},\mathbf{S}}(e) = * \ ].
\end{aligned}$$

Next, we check

$$\begin{aligned}
\mathbb{P}[\ \boldsymbol{\varrho}_d(e)=1\ ] &= \mathbb{P}[\ (\boldsymbol{\varrho}_0(e)=1) \vee \bigvee_{i=1}^{d} (\boldsymbol{\varrho}_{i-1}(e)=* \text{ and } \boldsymbol{\varrho}_i(x)=1)\ ] \\
&= \mathbb{P}[\ \boldsymbol{\varrho}_0(e)=1\ ] + \sum_{i=1}^{d} \mathbb{P}[\ \boldsymbol{\varrho}_{i-1}(e)=*\ ]\,\mathbb{P}[\ \boldsymbol{\varrho}_i(e)=1 \mid \boldsymbol{\varrho}_{i-1}(e)=*\ ] \\
&= q_e + \sum_{i=1}^{d} p_e (n^{-\delta/d})^{i-1} \frac{1-n^{-\delta/d}}{2} \\
&= q_e + p_e \frac{1-n^{-\delta}}{2}.
\end{aligned}$$

By setting $q_e := p_e(\frac{1}{2} + (1-p_e)n^{-\delta})$, we get

$$\mathbb{P}[\ \boldsymbol{\varrho}_d(e)=1\ ] = (1-p_e n^{-\delta})p_e = \mathbb{P}[\ R_{\mathbf{X},\mathbf{S}}(e)=1\ ].$$

(It is important to note that $q_e \approx \frac{1}{2}p_e$, so that $p_e + q_e \approx \frac{3}{2}p_e \leq 1$, which we require in order that $\boldsymbol{\varrho}_0$ is well-defined.[1])

We now complete the proof of Claim 15 as follows. For $i \in \{1,\ldots,d\}$, let Gates_i be the set of gates/inputs in the circuit C with distance $d-i$ from the output. In particular, Gates_d contains just the output gate of C, which computes the function f. We have

$$\begin{aligned}
\mathbb{P}[\ \mathcal{D}(f\restriction R_{\mathbf{X},\mathbf{S}}) > \delta \log n\ ] &= \mathbb{P}[\ \mathcal{D}(f\restriction \boldsymbol{\varrho}_d) > \delta \log n\ ] \\
&\leq \mathbb{P}\left[\ \bigvee_{i=1}^{d} \bigvee_{g\in\mathrm{Gates}_i} \left((\mathcal{D}(g\restriction \boldsymbol{\varrho}_i) > \delta\log n) \;\&\; \bigwedge_{h\in\mathrm{Gates}_{i-1}} (\mathcal{D}(h\restriction \boldsymbol{\varrho}_{i-1}) \leq \delta \log n) \right)\ \right] \\
&\leq \sum_{i=1}^{d} \sum_{g\in\mathrm{Gates}_i} \mathbb{P}\left[\ \mathcal{D}(g\restriction \boldsymbol{\varrho}_i) > \delta\log n \;\middle|\; \bigwedge_{h\in\mathrm{Gates}_{i-1}} (\mathcal{D}(h\restriction \boldsymbol{\varrho}_{i-1}) \leq \delta \log n)\ \right] \\
&\leq s \cdot (5n^{-\delta/d}\delta\log n)^{\delta\log n} \qquad \text{(by (4))}.
\end{aligned}$$

Since δ is a constant and $d = o(\frac{\log n}{\log\log n})$, we have $(5n^{-\delta/d}\delta\log n)^{\delta\log n} \leq n^{-\omega(\log\log n)}$ as follows:

$$\begin{aligned}
(5n^{-\delta/d}\delta\log n)^{\delta\log n} &\leq (n^{-\Omega(1/d)}\log n)^{\Omega(\log n)} \\
&= (n^{-\omega(\log\log n/\log n)}\log n)^{\Omega(\log n)} \\
&= (2^{-\omega(\log\log n)}\log n)^{\Omega(\log n)} \\
&= (\log n)^{-\omega(\log n)} \\
&= (n^{(\log\log n)/(\log n)})^{-\omega(\log n)} \\
&= n^{-\omega(\log\log n)}.
\end{aligned}$$

Since $s = n^{O(\log\log n)}$, we have $s\cdot(5n^{-\delta/d}\delta\log n)^{\delta\log n} = n^{-\omega(\log\log n)}$, which completes the proof. □

[1]Here we are assuming (effectively without loss of generality) that $\theta(e) > 0$ for all $e \in E(G)$. This is a minor technicality: in the case that $\theta(e) = 0$ for some $e \in E(G)$, there is a slightly different way to define restrictions $\boldsymbol{\varrho}_i$. However, we may ignore this case for all intents and purposes, since if $\theta(e) = 0$ then the average-case complexity of $\mathrm{SUB}(G)$ on $\mathbf{X}_\theta$ reduces to the average-case complexity of $\mathrm{SUB}(G-\{e\})$ on $\mathbf{X}_{\theta\restriction E(G)-\{e\}}$.