CSC2429, MAT1304: Circuit Complexity October 25, 2016

Lecture 7: The Subgraph Isomorphism Problem

Instructor: Benjamin Rossman

1 The Problem SUB(G)

Convention 1 (Graphs)**.**

Graphs are finite simple graphs with no isolated vertices. Formally, a graph G consists of a finite vertex set $V(G)$ and an edge set $E(G) \subseteq \binom{V(G)}{2}$ such that $\bigcup_{e \in E(G)} e = V(G)$. There are $2^{|E(G)|}$ subgraphs of G. (We may thus identify subgraphs G with elements of the Boolean hypercube $\{0,1\}^{E(G)}$.)

For $k \geq 2$: K_k denotes the complete graph of order k, and P_k denotes the path of order k.

Definition 2 (The Blow-Up $G^{\uparrow n}$ of a Graph G)**.**

For a graph G and $n \in \mathbb{N}$, let $G^{\uparrow n}$ denote the graph

$$V(G^{\uparrow n}) = \{v^{(i)} : v \in V(G),\ i \in [n]\},$$
$$E(G^{\uparrow n}) = \big\{\{v^{(i)}, w^{(j)}\} : \{v,w\} \in E(G),\ i,j \in [n]\big\}.$$

For $\alpha \in [n]^{V(G)}$, let $G^{(\alpha)}$ denote the graph

$$V(G^{(\alpha)}) = \big\{v^{(\alpha_v)} : v \in V(G)\big\},$$
$$E(G^{(\alpha)}) = \big\{\{v^{(\alpha_v)}, w^{(\alpha_w)}\} : \{v,w\} \in E(G)\big\}.$$

Each $G^{(\alpha)}$ is an isomorphic copy of G that sits inside $G^{\uparrow n}$.

We have a similar notation for subgraphs of G and their copies inside $G^{\uparrow n}$. For $H \subseteq G$ and $\beta \in [n]^{V(H)}$, let $H^{(\beta)}$ denote the graph

$$V(H^{(\beta)}) = \big\{v^{(\beta_v)} : v \in V(H)\big\},$$
$$E(H^{(\beta)}) = \big\{\{v^{(\beta_v)}, w^{(\beta_w)}\} : \{v,w\} \in E(H)\big\}.$$

For $X \subseteq G^{\uparrow n}$ and $H \subseteq G$, let

$$\mathrm{Sub}_H(X) = \big\{H^{(\beta)} : \beta \in [n]^{V(H)} \text{ such that } H^{(\beta)} \subseteq X\big\},$$
$$\mathrm{sub}_H(X) = |\mathrm{Sub}_H(X)|.$$

We refer to elements of $\mathrm{Sub}_H(X)$ as *H-subgraphs* of X.

Definition 3 (The Colored G-Subgraph Isomorphism Problem)**.**

For a graph G and $n \in \mathbb{N}$, let SUB(G, n) be the problem, given a subgraph $X \subseteq G^{\uparrow n}$, of determining whether or not there exists $\alpha \in [n]^{V(G)}$ such that $G^{(\alpha)} \subseteq X$. For complexity purposes, we regard SUB(G, n) as a Boolean function $\{0,1\}^{|E(G)| \cdot n^2} \to \{0,1\}$ with variables $\{X_e\}_{e \in E(G^{\uparrow n})}$. We refer to $\mathrm{SUB}(G) = \{\mathrm{SUB}(G,n)\}_{n \in \mathbb{N}}$ as the *colorful G-subgraph isomorphism problem.*

Note the brute-force upper bound: $\mathrm{SUB}(G)$ is solvable by monotone depth-2 (OR $\circ$ AND) formulas of size $O(|E(G)| \cdot n^{|V(G)|})$:

$$\bigvee_{\alpha \in [n]^{V(G)}} \bigwedge_{\{v,w\} \in E(G)} X_{\{v^{(\alpha v)}, w^{(\alpha w)}\}}.$$

As we will see later on, there is a better upper bound (for monotone circuits of $O(|V(G)|)$ depth) in terms of the tree-width of G. In the next few lectures, we will show nearly matching lower bounds for bounded-depth circuits, as well as monotone circuits.

Two important special cases of the problem $\mathrm{SUB}(G)$ are when $G = K_k$ (the "k-clique problem") and $G = P_k$ (the "distance-k connectivity problem").

2 Relationship to $\mathrm{SUB}_{\mathrm{uncolored}}(G)$

We remark on the relationship between $\mathrm{SUB}(G)$ and its uncolored version, denoted $\mathrm{SUB}_{\mathrm{uncolored}}(G)$. This is the problem: given a graph $X \subseteq K_n$ (i.e. a graph with $V(X) \subseteq \{1, \ldots, n\}$), determine whether or not X contains a subgraph isomorphic to G. (Note that k-CLIQUE is precisely the problem $\mathrm{SUB}_{\mathrm{uncolored}}(K_k)$.)

There is a well-known reduction from $\mathrm{SUB}_{\mathrm{uncolored}}(G)$ to $\mathrm{SUB}(G)$.

Theorem 4 ("Color-Coding Technique", Alon-Yuster-Zwick 1995)**.** *There is a quasi-linear size* AC^0 *reduction from* $\mathrm{SUB}_{\mathrm{uncolored}}(G)$ *to* $\mathrm{SUB}(G)$*.*

Proof Sketch. We are given an uncolored graph $X \subseteq K_n$ and wish to determine whether or not it contains a subgraph isomorphic to G. For a function $\varphi : [n] \to V(G)$, let $X^\varphi \subseteq G^{\uparrow n}$ be the graph with

$$E(X^\varphi) = \big\{\{v^{(i)}, w^{(j)}\} : \{v,w\} \in E(G) \text{ and } \{i,j\} \in E(X) \text{ and } \varphi(i) \neq \varphi(j)\big\}.$$

Suppose we have a family Φ of functions $\varphi : [n] \to V(G)$ with the property that, for every $U \subseteq [n]$ with $|U| = |V(G)|$, there exists $\varphi \in \Phi$ with $\varphi(U) = V(G)$. Such a family Φ is called a *k-perfect family of hash functions* where $k = |V(G)|$ and explicit constructions are known of size $O_k(\log n)$ (the constant here is exponential in k).

The reduction from $\mathrm{SUB}_{\mathrm{uncolored}}(G)$ to $\mathrm{SUB}(G)$ works as follows: For each $\varphi \in \Phi$, test whether $\mathrm{sub}_G(X^\varphi) \geq 1$. If $\mathrm{sub}_G(X^\varphi) \geq 1$ for any $\varphi \in \Phi$, then accept (i.e. conclude that X contain a subgraph isomorphic to G); otherwise, reject.

To see that this reduction is correct, note that if X does not contain a subgraph isomorphic to G, then clearly $\mathrm{sub}_G(X^\varphi) = 0$ for every function $\varphi : [n] \to V(G)$. On the other hand, if X contains a subgraph G' isomorphic to G, then for any $\varphi \in \Phi$ with $\varphi(V(G')) = V(G)$, we have $\mathrm{sub}_G(X^\varphi) \geq 1$.

The fact that this reduction can be implemented by AC^0 circuits of size $O(n \log n)$ was noted by Amano (2010). □

In many cases, there is a trivial reduction in the opposite direction from $\mathrm{SUB}(G)$ to $\mathrm{SUB}_{\mathrm{uncolored}}(G)$.

Definition 5. A graph G is a *core* if every homomorphism $G \to G$ is one-to-one (and hence an automorphism).

For example, the complete graph K_k is a core.

Lemma 6. *If G is a core and $X \subseteq G^{\uparrow n}$, then X contains a subgraph isomorphic to G if and only if* $\mathrm{sub}_G(X) \geq 1$.

Proof. In one direction, if $\mathrm{sub}_G(X) \geq 1$ then clearly X contains a subgraph isomorphic to G. In the other direction, assume X contains a subgraph G' isomorphic to G. Let $\varphi : V(G') \xrightarrow{\cong} V(G)$ be an isomorphism. Let $\pi : V(G^{\uparrow n}) \to V(G)$ be the homomorphism $v^{(i)} \mapsto v$. Then $\pi \circ \varphi^{-1}$ is a homomorphism $G \to G$ and, therefore, one-to-one (since G is a core). It follows that there exists an automorphism σ of G such that $\pi \circ \varphi^{-1} \circ \sigma : V(G) \to V(G)$ is the identity function. Then $G^{(\alpha)} \subseteq X$ where $\alpha \in [n]^{V(G)}$ is given by $\alpha_v = \varphi^{-1}(\sigma(v))$. □

Corollary 7. *If G is a core, then there is a linear-size AC^0 reduction (in fact, monotone projection) from $\mathrm{SUB}_{\mathrm{uncolored}}(G)$ to $\mathrm{SUB}(G)$.*

Proof. This reduction simply maps X regarded as an instance of $\mathrm{SUB}(G, n)$ (i.e. a subgraph $G^{\uparrow n}$) to X regarded as an instance of $\mathrm{SUB}_{\mathrm{uncolored}}(G, |V(G)| {\cdot} n)$ (i.e. a subgraph of $K_{|V(G)| \cdot n}$). □

Theorem 4 and Corollary 7 show that $\mathrm{SUB}(G)$ and $\mathrm{SUB}_{\mathrm{uncolored}}(G)$ are equivalent problems (up to quasi-linear AC^0 reductions) for cores G such as K_k. Henceforth, we shall focus almost entirely on the colored problem $\mathrm{SUB}(G)$, which appears to be more well-structured and better behaved than the uncolored version.

3 Threshold Weightings

For the uncolored subgraph isomorphism problem $\mathrm{SUB}_{\mathrm{uncolored}}(G)$, there is a canonical choice of input distribution (i.e. random graph) with respect to which the corresponding Boolean function $\{0,1\}^{\binom{n}{2}} \to \{0,1\}$ is balanced (i.e. has expectation bounded away from 0 and 1). This is the Erdos-Renyi random graph $\mathbf{G}(n, p)$ for an appropriate threshold value of p (see Example 15).

In the colored setting, we can identify a family of product distributions (what might be called "G-colored Erdos-Renyi random graphs") with respect to $\mathrm{SUB}(G)$ is balanced. This family is nicely indexed by a convex polytope of edge-weightings $E(G) \to [0, 2]$.

Definition 8. A *threshold weighting* on G is a function $\theta : E(G) \to \mathbb{R}_{\geq 0}$ such that

1. $\sum_{e \in E(H)} \theta(e) \leq |V(H)|$ for all $H \subseteq G$,
2. $\sum_{e \in E(G)} \theta(e) = |V(G)|$.

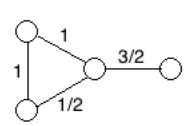
For $H \subseteq G$, we define

$$\Delta_\theta(H) := |V(H)| - \sum_{e \in E(H)} \theta(e).$$

Thus, condition (1) states that $\Delta_\theta(H) \geq 0$ for all $H \subseteq G$ and condition (2) states that $\Delta_\theta(G) = 0$.

We say that θ is *strict* if $\Delta_\theta(H) > 0$ for all $\emptyset \subset H \subset G$. (Note that θ is strict $\Rightarrow G$ is connected.)

Example 9.

1. Here is a threshold weighting on a particular 4-vertex graph: 

2. The set of threshold weightings on G is a polytope in $\mathbb{R}^{E(G)}$. In particular, if θ_1, θ_2 are threshold weightings, then so is every convex combination $\lambda\theta_1 + (1-\lambda)\theta_2$ for $0 < \lambda < 1$.

3. If G is r-regular, then the constant function $\theta \equiv 2/r$ is a threshold weighting.

   Exercise: If G is an r-regular expander, then $\Delta_{2/r}(H) \geq \Omega(\min\{|V(H)|, |V(G)| - |V(H)|\})$.

4. More generally, the function $\theta(\{v,w\}) := \frac{1}{\deg(v)} + \frac{1}{\deg(w)}$ is a threshold weighting.

4 The Random Graph $\mathbf{X}_\theta \subseteq G^{\uparrow n}$

Definition 10. For every threshold weighting θ on G, we define a random subgraph $\mathbf{X}_{\theta,n} \subseteq G^{\uparrow n}$ (we write $\mathbf{X}_\theta$ when n is understood from context) where, independently for all $\{v^{(i)}, w^{(j)}\} \in E(G^{\uparrow n})$,

$$\mathbb{P}\left[\ \{v^{(i)}, w^{(j)}\} \in E(\mathbf{X}_\theta)\ \right] = n^{-\theta(\{v,w\})}.$$

We will be interested in the average-case complexity of $\mathrm{SUB}(G)$ on $\mathbf{X}_\theta$.

Definition 11. If $f = (f_n)$ is a sequence of Boolean functions $f_n : \{\text{subgraphs of } G^{\uparrow n}\} \to \{0,1\}$, we say that *"$f$ solves $\mathrm{SUB}(G)$ in the average-case on $\mathbf{X}_\theta$"* if

$$\lim_{n\to\infty} \mathbb{P}\left[\ f_n(\mathbf{X}_{\theta,n}) = 1 \iff \mathrm{sub}_G(\mathbf{X}_{\theta,n}) \geq 1\ \right] = 1.$$

Clearly, the worst-case complexity of $\mathrm{SUB}(G)$ is lower-bounded by the maximum—over threshold weightings θ—of its average-case complexity with respect to $\mathbf{X}_\theta$. Our method for proving lower bounds on $\mathrm{SUB}(G)$ will consist of two parts:

(i) For each threshold weighting θ, we characterize the average-case AC^0 circuit size of $\mathrm{SUB}(G)$ on $\mathbf{X}_\theta$ in terms of a combinatorial parameter $\kappa_\theta(G)$ (defined in the next lecture): we show a lower bound of $n^{\kappa_\theta(G)-o(1)}$, as well as an upper bound of $n^{2\kappa_\theta(G)+O(1)}$.

(ii) For every graph G, we show that there exists a threshold weighting θ such that $\kappa_\theta(G) = \Omega(\mathbf{tw}(G)/\log \mathbf{tw}(G))$ where $\mathbf{tw}(G)$ is the tree-width of G. In special cases, such as $G = K_k$ or in general when G is an expander, we achieve optimal bound $\kappa_\theta(G) = \Omega(|V(G)|)$.

Together (i) and (ii) imply nearly tight lower bounds on the worst-case (really: worst average-case) AC^0 circuit size of of $\mathrm{SUB}(G)$.

The upper and lower bounds in (i) rely on properties of the random graph $\mathbf{X}_\theta$. In particular, we require a good understanding of the subgraph counts $\mathrm{sub}_H(\mathbf{X}_\theta)$ for $H \subseteq G$.

Lemma 12. *For all $H \subseteq G$, we have $\mathbb{E}\left[\ \mathrm{sub}_H(\mathbf{X}_\theta)\ \right] = n^{\Delta_\theta(H)}$. In particular, $\mathbb{E}\left[\ \mathrm{sub}_G(\mathbf{X}_\theta)\ \right] = 1$.*

Proof. Direct from definitions. □

In the case where θ is strict, we can say a lot more.

Theorem 13. *If θ is strict, then the following hold:*

1. $\mathrm{sub}_G(\mathbf{X}_\theta)$ *is asymptotically* Poisson(1). *That is, for every* $k \in \mathbb{N}$,

$$\lim_{n\to\infty} \mathbb{P}\left[\ \mathrm{sub}_G(\mathbf{X}_\theta) = k\ \right] = \frac{1}{ek!}.$$

   In particular, $\lim_{n\to\infty} \mathbb{P}[\ \mathrm{sub}_G(\mathbf{X}_\theta) \geq 1\ ] = 1 - \frac{1}{e}$ *(i.e.* SUB(G) *is balanced with respect to* $\mathbf{X}_\theta$*, justifying terminology "threshold weighting").*

2. *For all* $\emptyset \subset H \subset G$,

$$\begin{aligned} \mathbb{P}\left[\ \mathrm{sub}_H(\mathbf{X}_\theta) < \tfrac{1}{2} n^{\Delta_\theta(H)}\ \right] &\leq \exp\big(-\Omega(n^c)\big), \\ \mathbb{P}\left[\ \mathrm{sub}_H(\mathbf{X}_\theta) > \tfrac{3}{2} n^{\Delta_\theta(H)}\ \right] &\leq \exp\big(-\Omega(n^c)\big). \end{aligned}$$

   for a constant $c > 0$ *depending on* θ *and* H.

Time permitting, we will see elements of the proof of Theorem 13 in a future lecture.

5 More Examples of Threshold Weightings

Example 14 (Threshold weightings from Markov chains)**.** If $M \in V(G) \times V(G) \to [0,1]$ is the transition matrix for any Markov chain on G, that is,

1. $\sum_w M(v,w) = 1$ for every $v \in V(G)$ and
2. $M(v,w) > 0 \implies \{v,w\} \in E(G)$.

Then

$$\theta(\{v,w\}) := M(v,w) + M(w,v)$$

is a threshold weighting for G. In this case, we have

$$\Delta_\theta(H) = \sum_{(v,w)\,:\,v\in V(H) \text{ and } \{v,w\}\in E(G)\setminus E(H)} M(v,w),$$

that is, $\Delta_\theta(H)$ is equal to the amount of M-flow that leaves $V(H)$ via edges in $E(G) \setminus E(H)$. This is derived as follows:

$$\begin{aligned} \Delta_M(H) &= |V(H)| - \sum_{\{v,w\}\in E(H)} \Big(M(v,w) + M(w,v)\Big) \\ &= \sum_{v\in V(H)} \left(1 - \sum_{w\,:\,\{v,w\}\in E(H)} M(v,w)\right) \\ &= \sum_{v\in V(H)} \left(\sum_{w\,:\,\{v,w\}\notin E(H)} M(v,w)\right) \\ &= \sum_{(v,w)\,:\,v\in V(H) \text{ and } \{v,w\}\in E(G)\setminus E(H)} M(v,w). \end{aligned}$$

Example 15. The *threshold exponent* of G is defined by

$$\tau(G) := \min_{H \subseteq G} \frac{|V(H)|}{|E(H)|}.$$

This constant characterizes the appearance of subgraphs isomorphic to G in the Erdos-Renyi random graph $\mathbf{G}(n,p)$: for $p(n) = n^{-\tau(G)}$, the probability that $\mathbf{G}(n,p)$ contains a subgraph isomorphic to G is bounded away from 0 and 1 (while this probability tends to 0 or 1 for $p \ll n^{-\tau(G)}$ and $p \gg n^{-\tau(G)}$ respectively). The Erdos-Renyi random graph $\mathbf{G}(n, n^{-\tau(G)})$ is a canonical input distribution for studying the average-case complexity of $\mathrm{SUB}_{\mathrm{uncolored}}(G)$.

The average-case analysis of $\mathrm{SUB}_{\mathrm{uncolored}}(G)$ on $\mathbf{G}(n, n^{-\theta(G)})$ is essentially equivalent to the average-case analysis of $\mathrm{SUB}(G)$ on $\mathbf{X}_\theta$ for the threshold weighting $\theta : E(G) \to [0,2]$ defined by

$$\theta(\{v,w\}) := \begin{cases} \tau(G) & \text{if } \{v,w\} \in E(H) \text{ for any } H \subseteq G \text{ with } \tau(G) = \dfrac{|V(H)|}{|E(H)|}, \\ 0 & \text{otherwise.} \end{cases}$$

All of the lower/upper bounds we show for average-case $\mathrm{SUB}(G)$ on $\mathbf{X}_\theta$ carry over to $\mathrm{SUB}_{\mathrm{uncolored}}(G)$ on $\mathbf{G}(n, n^{-\theta(G)})$ for this particular setting of θ. (This is not by a formal reduction; I am only claiming that the calculations work out similarly.) We get more powerful results by working with the colored version $\mathrm{SUB}(G)$, thanks to the freedom to choose optimal θ. For this reason, going forward we will focus entirely on the average-case complexity of $\mathrm{SUB}(G)$ on $\mathbf{X}_\theta$ and leave the uncolored setting as a special case.

Remark 16. G is *balanced* if $\tau(G) = |V(G)|/|E(G)|$, and G is *strictly balanced* if $\tau(G) < |V(H)|/|E(H)|$ for all $H \subset G$. If G is balanced, then the above-defined θ is identically $\tau(G)$. If G is strictly balanced, then this θ is strict.