CS271 Randomness & Computation **Spring 2020**

Lecture 4: January 30

Instructor: Alistair Sinclair

4.1 A deterministic algorithm for primality testing

We conclude our discussion of primality testing by sketching the route to the first polynomial time deterministic primality testing algorithm announced in 2002. This is based on another randomized algorithm due to Agrawal and Biswas [AB99], which was subsequently derandomized by Agrawal, Kayal and Saxena [AKS02]. We won't discuss the derandomization in detail as that is not the main focus of the class.

The Agrawal-Biswas algorithm exploits a different number theoretic fact, which is a generalization of Fermat's Theorem, to find a witness for a composite number. Namely:

Fact 4.1 *For every $a > 1$ such that $gcd(a, n) = 1$, n is a prime iff $(x - a)^n = x^n - a \mod n$.*

Exercise: Prove this fact. [Hint: Use the fact that $\binom{n}{k} = 0 \bmod n$ for all $0 < k < n$ iff n is prime.]

The obvious approach for designing an algorithm around this fact is to use the Schwartz-Zippel test to see if $(x - a)^n - (x^n - a)$ is the zero polynomial. However, this fails for two reasons. The first is that if n is not prime then $\mathbf{Z}_n$ is not a field (which we assumed in our analysis of Schwartz-Zippel); the second is that the degree of the polynomial is n, which is the same as the cardinality of $\mathbf{Z}_n$ and thus too large (recall that Schwartz-Zippel requires that values be chosen from a set of size strictly larger than the degree).

Instead, we will use a variant of fingerprinting. Rather than a prime divisor as in standard fingerprinting, in this case our witness will be a low-degree polynomial, $r(x)$, such that $(x + 1)^n \neq x^n + 1 \mod (r(x), n)$. Plainly, if such a polynomial $r(x)$ exists then, from the above Fact, we can be sure that n is composite; while if n is prime then no such $r(x)$ can exist. Thus we will again get an algorithm with one-sided error. As usual, the challenge is to show that the density of witnesses is large enough if we pick them from a suitable set.

Here is the algorithm:

if n has a divisor less than 17 **or if** n is a perfect power **then output** "composite"
let $d = \lceil \log n \rceil$
let $r(x) = x^d + r_{d-1}x^{d-1} + \cdots + r_1 x + r_0$ where each r_i is chosen from $\mathbf{Z}_n$ uniformly at random
if $(x + 1)^n \neq x^n + 1 \mod (r(x), n)$ **then**
    output "composite"
else
    output "prime?"

Note that this algorithm can be implemented to run in time polylog(n), since all arithmetic is done modulo the polynomial $r(x)$ (of degree $\log n$) with coefficients in $\mathbf{Z}_n$, and the power $(x + 1)^n$ can be computed by repeated squaring using $O(\log n)$ multiplications.

4.1.1 The likelihood of finding a witness

For the analysis, let us assume that n is composite. By the first line of the algorithm, we may also assume that n is not a prime power, and has a prime divisor $p \geq 17$. Let $Q(x) = (x+1)^n - (x^n+1)$. We claim first that not only do we have $Q(x) \neq 0 \bmod n$, but also $Q(x) \neq 0 \bmod p$. (**Exercise:** prove this, along similar lines to the proof of the Fact above.) Now we can argue over $\mathbf{Z}_p$ instead of $\mathbf{Z}_n$. Since $Q(x) \neq 0$ over $\mathbf{Z}_p$, it has a unique factorization (up to a constant). Also, since $r(x)$ is a random (monic) polynomial over $\mathbf{Z}_n$ it is also a random (monic) polynomial over $\mathbf{Z}_p$. And, as $Q(x)$ has degree n, it can have at most $\frac{n}{d}$ irreducible factors of degree d.

Now, we can discuss the probability that $r(x)$ is a witness to n being a composite number, i.e., that $r(x)$ is not a factor of $Q(x) \pmod p$. To do this we will restrict attention to irreducible polynomials $r(x)$, since we don't have good control over the number of non-irreducible factors of $Q(x)$. Thus we have

$$\begin{aligned}
\Pr[r(x) \text{ is a witness}] &\geq \Pr[(r(x) \text{ is an irreducible witness}] \\
&\geq \Pr[r(x) \text{ is irreducible and } r(x) \text{ is not a factor of } Q(x)] \\
&= \Pr[r(x) \text{ is irreducible}] - \Pr[r(x) \text{ is an irreducible factor of } Q(x)].
\end{aligned}$$

A known fact is that the number of irreducible monic polynomials of degree d over $\mathbf{Z}_p$ is at least $\frac{p^d}{d} - \sqrt{p^d}$. When $p \geq 17$ and $d > 4$ this is at least $\frac{p^d}{2d}$. This allows us to lower bound the probability that $r(x)$ is irreducible:

$$\Pr[r(x) \text{ is irreducible}] \geq \frac{p^d/2d}{p^d} = \frac{1}{2d}.$$

Similarly, we can bound the probabilty that $r(x)$ divides $Q(x)$.

$$\Pr[r(x) \text{ is an irreducible factor of } Q(x)] \leq \frac{n/d}{p^d} \leq \frac{1}{4d}.$$

The last inequality is due (very crudely) to the fact that $p \geq 17$ and $d = \lceil \log n \rceil$.

Putting these together, we have that

$$\Pr[r(x) \text{ is a witness}] \geq \frac{1}{2d} - \frac{1}{4d} = \frac{1}{4d} = \Omega(\frac{1}{\log n}).$$

Thus when n is composite the algorithm finds a witness with probability $\Omega(\frac{1}{\log n})$. This means that $O(\log n)$ repeated trials suffice to find a witness with high probability.

4.1.2 Derandomization

The deterministic algorithm of [AKS02] is based on a derandomization of the above algorithm. The (very rough) idea is to use a *fixed* polynomial $r(x) = x^d - 1$ for $d \approx (\log n)^2$, and to search instead through the values of a, testing whether $(x-a)^n \neq x^n - a \bmod (r(x), n)$. It turns out that checking each value of a up to about $(\log n)^2$ is guaranteed to reveal a witness if n is composite. The running time of the resulting algorithm is roughly $O((\log n)^6)$, making it not useful in practice. We omit the details.

4.2 The Probabilistic Method

The probabilistic method is a powerful mathematical tool used in solving many combinatorial problems. The method was first introduced by Paul Erdős in his seminal 1947 paper [E47].

The probabilistic method works as follows: to show that an object with a specific property exists, we choose an object randomly from an appropriate collection of objects and show that it has the desired property with positive probability, which proves the existence of at least one such object. It turns out that it is often easier to show the existence of certain objects by this method than by constructing them explicitly. Moreover, the probabilistic method often suggests efficient randomized algorithms, which can sometimes subsequently be derandomized.

4.2.1 Ramsey Theory

As a warm-up, we state and prove a result from Erdős's original paper on Ramsey theory [E47].

Definition: The k-th (diagonal) Ramsey number, R_k, is the smallest number n such that any 2-coloring of the edges of the complete graph on n vertices, K_n, must contain a monochromatic k-clique.

As a useful **Exercise**, you should verify that $R_3 = 6$ by showing that K_5 can be 2-colored with no monochromatic triangles, while K_6 cannot.

The existence of Ramsey numbers can be established by a simple inductive argument, which also provides an upper bound of 2^{2k-3} for R_k. The following theorem provides a lower bound for R_k via the probabilistic method.

Theorem 4.2 $R_k > 2^{k/2}$.

Proof: Let $n = 2^{k/2}$. We show that there exists a 2-coloring of the complete graph K_n, such that there is no monochromatic k-clique. Randomly color each edge of K_n red or blue with probability $\frac{1}{2}$, independently. Let C be any k-clique in the graph K_n. Then,

$$\Pr[C \text{ is monochromatic}] = 2 \cdot 2^{-\binom{k}{2}} = 2^{1-\binom{k}{2}}.$$

Since the total number of k-cliques in K_n is $\binom{n}{k}$, by the union bound, we get

$$\begin{aligned} \Pr[K_n \text{ has a monochromatic } k\text{-clique}] &\leq \binom{n}{k} \cdot \Pr[\text{a given } k\text{-clique is monochromatic}] \\ &= \binom{n}{k} \cdot 2^{1-\binom{k}{2}} \\ &\leq \frac{n^k}{k!} \cdot 2^{1-\binom{k}{2}}. \end{aligned}$$

Putting $n = 2^{k/2}$, we get

$$\begin{aligned} \Pr[K_n \text{ has a monochromatic } k\text{-clique}] &\leq \frac{1}{k!} \cdot 2^{\frac{k^2}{2}+1-\frac{k^2-k}{2}} \\ &= \frac{1}{k!} \cdot 2^{\frac{k+2}{2}} \\ &< 1 \quad \text{for } k \geq 3. \end{aligned}$$

So, the probability of a good coloring, i.e., a 2-coloring such that there is no monochromatic k-clique, is more than zero. Thus, there must exist a good coloring. ■

Ramsey numbers have been the object of extensive study in combinatorics. Despite this, improvements in bounds on R_k have been slow. In fact, for general k the above upper and lower bounds remain essentially the

best known, in the sense that no asymptotic lower bound of the form $R_k \geq 2^{(1/2+\epsilon)k}$ or upper bound of the form $R_k \leq 2^{(2-\epsilon)k}$ for $\epsilon > 0$ has been found. In particular, it is surprising that a completely random coloring of K_n provides a lower bound which is almost as good as bounds obtained by the cleverest deterministic arguments.

4.2.2 The Max Cut Problem

We now consider the problem of finding a partition of a graph into two parts such that the number of cut edges (i.e., those between the two parts of the partition) is maximized.

Input: A graph $G = (V, E)$.

Goal: Find a partition $V = V_1 \cup V_2$ of G such that the number of edges between V_1 and V_2 is maximized.

This is a well-known NP-hard optimization problem. However, as the following lemma shows, a simple application of the probabilistic method gives us a lower bound on the size of the maximum cut.

Lemma 4.3 *There exists a cut containing at least $\frac{|E|}{2}$ edges.*

Proof: Pick a random partition, by assigning each vertex to V_1 or V_2 independently with probability $\frac{1}{2}$. For each edge e, define the indicator variable X_e as

$$X_e = \begin{cases} 1 & \text{if } e \text{ is cut;} \\ 0 & \text{otherwise} \end{cases}$$

Thus, if X is the number of cut edges, then $X = \sum_{e \in E} X_e$. Moreover, we have

$$\mathrm{E}[X_e] = \Pr[e \text{ is cut}] = \Pr[\text{the endpoints of } e \text{ lie on different sides of the partition}] = \frac{1}{2}.$$

Now by linearity of expectation,

$$\mathrm{E}[X] = \sum_{e \in E} \mathrm{E}[X_e] = \frac{|E|}{2}.$$

This implies that there must exist a cut such that $X \geq \frac{|E|}{2}$ (since any random variable must achieve a value at least as large as its expectation at some sample point). ■

Exercise: Improve the above bound slightly by showing that there must always exist a cut with *strictly more than* $\frac{|E|}{2}$ edges.

4.2.3 Independent Set

In a graph $G = (V, E)$, a subset of vertices $V' \subset V$ is said to be an *independent set* if no two vertices $u, v \in V'$ are adjacent in G. It is NP-hard to determine the size of a largest independent set in G. However, an elementary but clever application of the probabilistic method will allow us to establish a good lower bound on the size of the maximum independent set for any given graph.

Claim 4.4 *Any graph $G = (V, E)$ contains an independent set $V' \subset V$ such that $|V'| \geq \sum_{v \in V} \frac{1}{\deg(v)+1}$, where $\deg(v)$ is the number of vertices adjacent to v in G.*

As an example, this claim implies that every 3-regular graph contains an independent set of size at least $|V|/4$.

Proof: Assign a random weight w_v to every vertex $v \in V$, choosing these weights independently and uniformly at random from the interval $[0, 1]$. Call $v \in V$ a *local minimum* if $w_v < w_u$ for every vertex u adjacent to v. Since no two adjacent vertices can both be local minima, it is clear that the set of local minima forms an independent set. Moreover, since weights are assigned uniformly from a continuous set, any vertex among v and its neighbors is equally likely to have minimum weight, so we have

$$\Pr[v \text{ is a local minimum}] = \frac{1}{\deg(v) + 1}.$$

Now let X be the number of local minima, and write $X = \sum_v X_v$, where X_v is an indicator r.v. defined by

$$X_v = \begin{cases} 1 & \text{if } v \text{ is a local minimum;} \\ 0 & \text{otherwise.} \end{cases}$$

By linearity of expectation,

$$\mathbf{E}[X] = \sum_{v \in V} \mathbf{E}[X_v] = \sum_{v \in V} \frac{1}{\deg(v) + 1}.$$

Hence G must contain an independent set of the claimed size. ■

4.2.4 Graph Crossing Number

Consider an arbitrary graph $G = (V, E)$; let n denote the number of vertices in G and let m denote the number of edges. We define the *crossing number* $c(G)$ as the minimum number of edge crossings in any planar embedding of G. (Thus a graph is planar iff $c(G) = 0$.)

Most graphs cannot be embedded in the plane without crossing edges. In fact, those that can are necessarily quite sparse. Euler's formula says that if G is planar then

$$m \leq 3n - 6. \tag{4.1}$$

We begin with a preliminary lower bound on the crossing number that generalizes Euler's formula:

Claim 4.5 *For any graph G with n vertices and m edges, we have $c(G) \geq m - 3n + 6$.*

Proof: The proof is a simple deterministic argument.

Consider an optimal planar embedding of G, i.e., an embedding that achieves $c = c(G)$ edge crossings. It is easy to verify (**Exercise**) that such an embedding satisfies the following properties:

1. No edge crosses itself.
2. No two edges cross more than once.
3. No two edges that share a vertex cross.

Construct a new graph $G' = (V', E')$ from G by replacing each edge crossing in the optimal embedding of G with a vertex, as shown in Figure 4.1. More precisely, we start with $G' = G$ and, for every pair of edges $e_1 = (v_a, v_b)$ and $e_2 = (v_c, v_d)$ that cross, transform G' as follows:

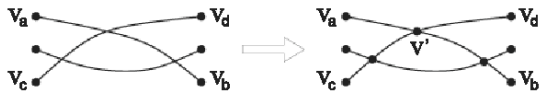

Figure 4.1: Construction of planar graph G' from G.

1. Add a new vertex v' to V'.
2. Remove e_1 and e_2 from E' and insert four new edges: $e'_1 = (v_a, v')$, $e'_2 = (v_b, v')$, $e'_3 = (v_c, v')$, and $e'_4 = (v_d, v')$.

From the construction and the properties of optimal planar embeddings, we have: $|V'| \equiv n' = n + c$ and $|E'| \equiv m' = m + 2c$. (To verify the second equality, observe that the number of edges increases by two for every edge crossing eliminated from the original graph). Furthermore, G' is a planar graph, so by Euler's formula we have:

$$\begin{aligned} m' &\leq 3n' - 6 \\ m + 2c &\leq 3n + 3c - 6 \\ c &\geq m - 3n + 6, \end{aligned}$$

as required. ■

The above result gives us an estimate of the crossing number that is reasonably tight for sparse graphs (when m is not much larger than $3n$). We now apply the probabilistic method to demonstrate a much stronger lower bound for large values of m. This beautiful proof is attributed to Chazelle, Sharir and Welzl [CSW], though the result (with a different constant) had previously been proved using much heavier deterministic tools in [L83] and [ACNS82].

Claim 4.6 *For any graph G with n vertices and m edges with $m \geq 4n$, we have $c(G) \geq \frac{m^3}{64n^2}$.*

Proof: Fix an optimal planar embedding of G with $c = c(G)$ crossings. Construct a *random induced subgraph* G_p of G, by including each vertex in G_p independently with probability p (and retaining precisely those edges both of whose endpoints are in G_p). The value of p will be specified later.

Now let n_p and m_p be random variables denoting the number of vertices and edges in G_p, respectively. Additionally, let c_p denote the number of edge crossings from the original embedding that remain in G_p. Using the result of Claim 5.4, we have:

$$c_p \geq m_p - 3n_p + 6 \geq m_p - 3n_p.$$

Taking expectations then gives

$$\mathbf{E}[c_p] \geq \mathbf{E}[m_p - 3n_p] = \mathbf{E}[m_p] - 3\mathbf{E}[n_p].$$

Now, since each vertex in G is included in G_p with probability p, we have $\mathbf{E}[n_p] = np$. Similarly, since every edge in G_p has probability p^2 of remaining in G_p, we have $\mathbf{E}[m_p] = mp^2$. Lastly, $\mathbf{E}[c_p] = cp^4$, since every edge crossing involves four distinct vertices of G. Plugging these values into the above inequality gives

$$cp^4 \geq mp^2 - 3pn,$$

and hence

$$c \geq \frac{m}{p^2} - \frac{3n}{p^3}.$$

Now we choose p to maximize the right-hand side of this inequality. Specifically, setting $p = 4n/m$ yields

$$c \geq \frac{m^3}{64n^2},$$

as claimed. (A slightly better constant is obtained by setting $p = 9n/2m$, but at the cost of requiring $m \geq (9/2)n$.) ■

Note that the bound proved in Claim 4.6 already beats the vanilla bound in Claim 4.5 when $m = 6n$.

References

[AB99] M. Agrawal and S. Biswas, "Primality and identity testing via Chinese remaindering," *Proceedings of IEEE FOCS* 1999, pp. 202–209. Full version appeared in *Journal of the ACM* **50** (2003), pp. 429–443.

[AKS02] M. Agrawal, N. Kayal and N. Saxena, "PRIMES is in P," *Annals of Mathematics* **160** (2004), pp. 781–793. Result first announced in a preprint, August 2002: http://www.cse.iitk.ac.in/users/manindra/primality_original.pdf

[ACNS82] M. Ajtai, V. Chvátal, M. Newborn and E. Szemerédi, "Crossing-free subgraphs," *Annals of Discrete Mathematics* **12** (1982), pp. 9–12.

[CSW] B. Chazelle, M. Sharir and E. Welzl, folklore.

[E47] P. Erdős, "Some remarks on the theory of graphs," *Bulletin of the American Math. Society* **53** (1947), pp. 292–294.

[L83] F.T. Leighton, *Complexity issues in VLSI*, MIT Press, Cambridge, 1983.