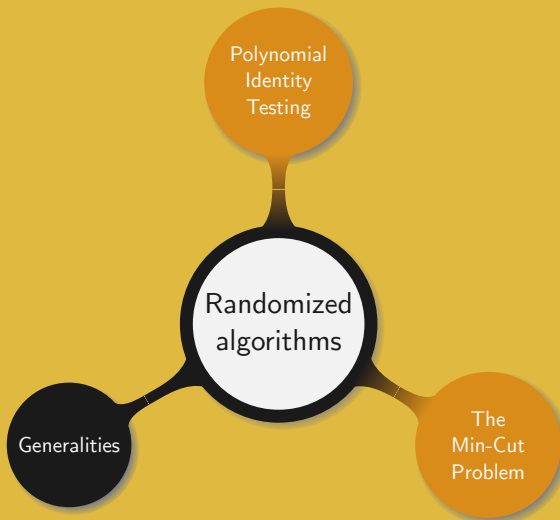


```
1001011111010100010101001111000111000100010111000000100110100001111101
1010010110011100111011101010001110000010001101110101011011111101010000
11011100111000110100110101100110110101010100111101000010101010101000
000110011110001000000110101011011110111100110000001010101011110100001
1100101110000100001001010110000011100010101000100110110010001101100100
010101101011110101001100101110100111100101001111001011101001110100
0110001111101000011010011111010111001100110111001100110111101001101
1110100100000100111101111010010001100110011001100110011001100110011001
10011110001111010001010111010101010101010101010101010101010101010101
0111110111111010011001001001001001001001001001001001001001001001001001
111011011001111000010010101010101010101010101010101010101010101010101
1101101000100110010100100111100001010011100001000101110010011011010011
0000110101010001100110000111000100101110011110000101001101100100000110
110111101001010101110101111000010100010011101100000010110111100000011
011110100010010101101100111001110101101110010110101011001100110101100
1011011011011111010000000011101110001111010111011001110011010011100000
```

Introduction to Algorithms

5. Randomized algorithms

Manuel – Fall 2021



Along chapter 2, section 3 (2.38) we studied:

- How to model computers
- What an algorithm is
- Various classes of problems

Along chapter 2, section 3 (2.38) we studied:

- How to model computers
- What an algorithm is
- Various classes of problems

Computation on a deterministic Turing machine depends on:

- The state of the machine
- The symbol read on the tape
- In any situation at most one action is performed

Computation on a non-deterministic Turing machine:

- The machine has a state and a symbol is read on the tape
- The machine branches into many copies
- The machine transitions into one of the copies

When the machine is run more than once:

- Different paths are chosen
- Computation cannot be exactly reproduced

Computation on a non-deterministic Turing machine:

- The machine has a state and a symbol is read on the tape
- The machine branches into many copies
- The machine transitions into one of the copies

When the machine is run more than once:

- Different paths are chosen
- Computation cannot be exactly reproduced

Remark. A non-deterministic Turing machine can be simulated by a deterministic Turing machine with three tapes.

Two main types of algorithms:

- Deterministic: a fixed sequence of steps if followed and the output is completely determined by the input
- Probabilistic: add some randomness to the process
 - Monte Carlo algorithm: returns either True or unknown; increasing the number of test decreases the probability of having a “false positive”
 - Las Vegas algorithms: always returns a correct result; the running time is however random; the time complexity cannot be precisely evaluated

Quick sort:

- Worst case: $\mathcal{O}(n^2)$ vs. average case $\mathcal{O}(n \log n)$
- Fixed pivot: if the list to sort is originally structured worst complexity is likely to apply
- Random pivot: even if the list to sort is originally structured choosing a random pivot is equivalent to sorting with a fixed pivot on a randomly ordered list
- Running time: random depending on the choice of the pivot

Primality testing:

- AKS: deterministic in time $\tilde{O}(n \log^{12} n)$
- **Miller-Rabin:**
 - Complexity: $\tilde{O}(k \log^2 n)$, with k the number of witnesses
 - Returns composite or unknown
 - Unknown means prime with probability $1 - 4^{-k}$

Definitions

- ① A non-deterministic Turing machine which chooses a random transitions according to some probability distribution is called a *probabilistic Turing machine*.
- ② A language L is in the Bounded-error Probabilistic Polynomial time complexity class (BPP) if and only if there is a probabilistic Turing machine M such that
 - For any input, M runs in polynomial time;
 - For all x in L , M returns 1 with probability larger than $1/2 + \epsilon$, with $\epsilon > 0$;
 - For all x not in L , M returns 1 with probability less than $1/2 - \epsilon$, with $\epsilon > 0$;

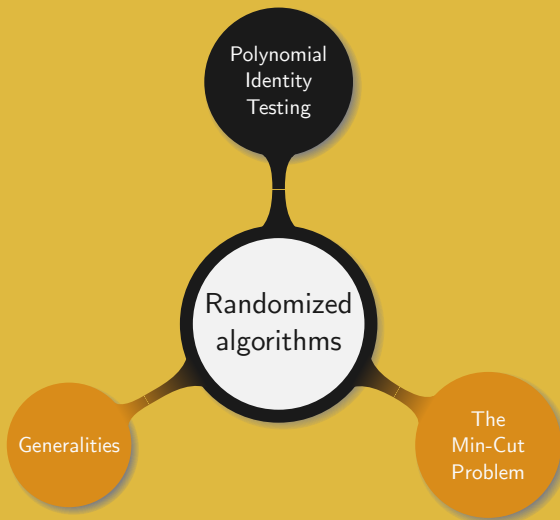
Remark. In practice instead of taking $1/2 + \epsilon$, $2/3$ and $1/3$ are often chosen.

Reasons for using probabilistic algorithms:

- No polynomial time deterministic algorithm is known
- Available polynomial time deterministic algorithms are slower

Open questions:

- Is $\mathcal{P} = \text{BPP}$?
- Are probabilistic algorithm more powerful than deterministic ones?
- What is the relative power of probabilistic and deterministic computations?



Problem (Polynomial Identity Testing (PIT))

Let P be a multivariate polynomial over some field. Decide whether P is identically zero.

Remark. Identically zero means that after expanding P it reduces to zero. In particular this is a different problem from the Evaluate to Zero Everywhere (EZE) problem where one wants to decide if an n -multivariate polynomial evaluates to zero for all x_i , $1 \leq i \leq n$.

The univariate case can be easily solved:

- The polynomial is given as a sum of monomials:
 - Reduce the sum
 - Check all the monomials and return True if they all equal zero
- The polynomial P of degree d is in a more complex form:
 - Arbitrarily choose $d + 1$ points (e.g. $0, \dots, d$)
 - Evaluate P at those $d + 1$ points
 - Conclude that $P = 0$ if and only if they all evaluate to zero

The univariate case can be easily solved:

- The polynomial is given as a sum of monomials:
 - Reduce the sum
 - Check all the monomials and return True if they all equal zero
- The polynomial P of degree d is in a more complex form:
 - Arbitrarily choose $d + 1$ points (e.g. $0, \dots, d$)
 - Evaluate P at those $d + 1$ points
 - Conclude that $P = 0$ if and only if they all evaluate to zero

What is the issue with multivariate polynomials?

Definitions

- ① A *monomial* is an expression of the form $\alpha \prod_{i=1}^n x_i^{\beta_i}$, where α is an element from a base field, the x_i are n variables, and the β_i are positive integers.
- ② The *total degree of a monomial* is $\sum_i \beta_i$.
- ③ The *total degree of a polynomial* is the largest total degree among all the monomials composing the polynomial.

Definitions

- ① A *monomial* is an expression of the form $\alpha \prod_{i=1}^n x_i^{\beta_i}$, where α is an element from a base field, the x_i are n variables, and the β_i are positive integers.
- ② The *total degree of a monomial* is $\sum_i \beta_i$.
- ③ The *total degree of a polynomial* is the largest total degree among all the monomials composing the polynomial.

Lemma (Schwartz-Zippel)

Let P be a n -multivariate polynomial of total degree d , that is not identically zero, over a field \mathbb{F} . For y_1, \dots, y_n , chosen uniformly and independently from a finite set $S \subset \mathbb{F}$,

$$\Pr [P(y_1, \dots, y_n) = 0] \leq \frac{d}{|S|}.$$

Proof. We proceed by induction on the number of variables n .

Base case: For $n = 1$ the result is clear as a univariate polynomial of degree d has at most d roots.

Induction step: we assume that the result is true for an $(n-1)$ -multivariate polynomial and prove it is also true in the case of n variables.

Let k be the largest power of X_1 in any monomial composing P . Then

$$P(X_1, \dots, X_n) = \sum_{i=0}^k X_1^i Q_i(X_2, \dots, X_n).$$

By construction, Q_k is not identically zero, its total degree is at most $d-k$, and it has $n-1$ variables. Therefore by the induction hypothesis we get

$$\Pr [Q_k(y_2, \dots, y_n) = 0] \leq \frac{d-k}{|S|}.$$

For y_2, \dots, y_n in \mathbb{F} , we call \mathcal{E}_1 the event $Q_k(y_2, \dots, y_n) = 0$.

Selecting y_2, \dots, y_n such that \mathcal{E}_1 does not occur, we define $R(X_1)$ to be the polynomial

$$R(X_1) = \sum_{i=0}^k X_1^i Q_i(y_2, \dots, y_n) = P(X_1, y_2, \dots, y_n).$$

Clearly $R(X_1)$ is not identically zero since \mathcal{E}_1 did not occur, meaning that X_1^k has a non zero coefficient. Therefore

$$\Pr [R(y_1) = 0 \mid \neg \mathcal{E}_1] \leq \frac{k}{|S|}.$$

Let \mathcal{E}_2 be the event $R(y_1) = 0$, which can also be stated as $P(y_1, \dots, y_n) = 0$. In order to prove the lemma it remains to bound $\Pr [\mathcal{E}_2]$.

As we have already bounded $\Pr[\mathcal{E}_2 \mid \neg \mathcal{E}_1]$ and $\Pr[\mathcal{E}_1]$ we can rewrite $\Pr[\mathcal{E}_2]$ as

$$\begin{aligned}\Pr[\mathcal{E}_2] &= \Pr[\mathcal{E}_2 \wedge \mathcal{E}_1] + \Pr[\mathcal{E}_2 \wedge \neg \mathcal{E}_1] \\ &= \Pr[\mathcal{E}_2 \wedge \mathcal{E}_1] + \Pr[\mathcal{E}_2 \mid \neg \mathcal{E}_1] \Pr[\neg \mathcal{E}_1] \\ &\leq \Pr[\mathcal{E}_1] + \Pr[\mathcal{E}_2 \mid \neg \mathcal{E}_1] \\ &\leq \frac{d-k}{|S|} + \frac{k}{|S|} = \frac{d}{|S|}\end{aligned}$$

Hence, by the induction principle Schwartz-Zippel lemma holds. □

Remark. Schwartz-Zippel lemma (5.13) says that when evaluating P at a random point there is a very low probability of finding a root. It however does not mean that a polynomial over \mathbb{R} has finitely many roots. For instance $P(X_1, X_2) = X_1$ has infinitely many roots.

Algorithm.

Input : $P(X_1, \dots, X_n)$ of degree d and a field \mathbb{F} , l and k

Output: not zero or probably zero

- 1 Select a subset $S \subset \mathbb{F}$ of size $\geq ld$;
 - 2 **for** $i \leftarrow 1$ **to** k **do**
 - 3 $(y_1, \dots, y_n) \leftarrow \text{rand}(S)$;
 - 4 **if** $P(y_1, \dots, y_n) \neq 0$ **then return** not zero;
 - 5 **end for**
 - 6 **return** probably zero;
-

Algorithm.

Input : $P(X_1, \dots, X_n)$ of degree d and a field \mathbb{F} , l and k

Output: not zero or probably zero

```
1 Select a subset  $S \subset \mathbb{F}$  of size  $\geq ld$ ;  
2 for  $i \leftarrow 1$  to  $k$  do  
3    $(y_1, \dots, y_n) \leftarrow \text{rand}(S)$ ;  
4   if  $P(y_1, \dots, y_n) \neq 0$  then return not zero;  
5 end for  
6 return probably zero;
```

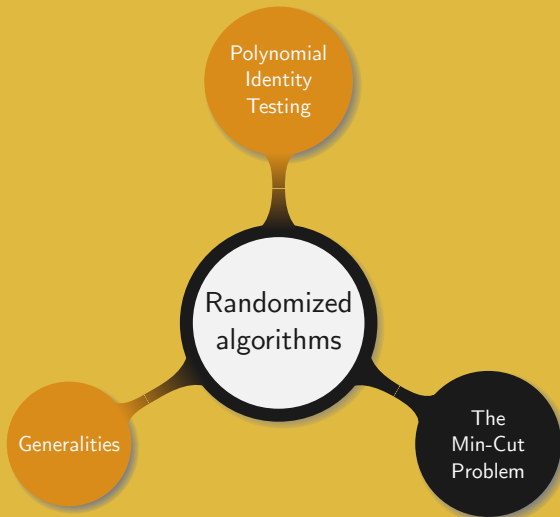
Theorem

Let $P(X_1, \dots, X_n)$ be a polynomial. If $P(X_1, \dots, X_n)$ is not identically zero, then algorithm 5.17 returns the correct result with probability at least $1 - 1/l^k$.

Proof. For each iteration of the loop the probability of testing a root is at most $1/l$. Each choice of element in S is independent of the previous ones, such that the probability of passing all the tests while $P(X_1, \dots, X_n)$ is not zero is at most $1/l^k$.

Hence the probability of returning “probably zero” while $P(X_1, \dots, X_n)$ is not identically zero is at least $1 - 1/l^k$. \square

Remark. If the field \mathbb{F} has less elements than the number of roots of $P(X_1, \dots, X_n)$ then Schwartz-Zippel lemma (5.13) is of no use. It is however possible to overcome this problem by applying algorithm 5.17 to the polynomial $P(X_1, \dots, X_n)$ over an extension field of \mathbb{F} .



Problem (Min-Cut)

Given a connected multi-graph G , determine the minimum number of edges that must be removed such that G becomes disconnected.

Common applications:

- Split a problem for parallel programming
- Optimize a divide and conquer strategy
- Segment an image into regions of similar color/texture

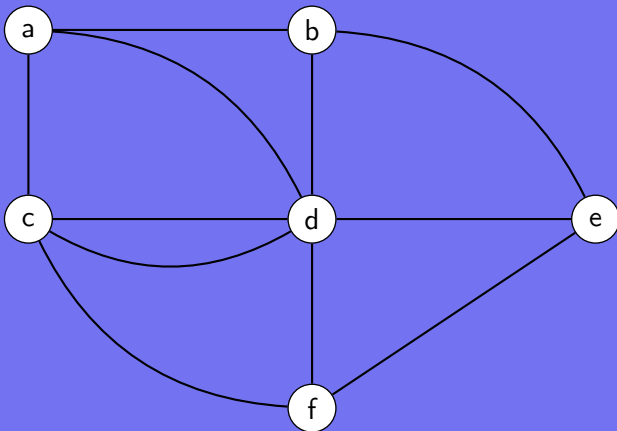
Algorithm. (*Karger*)

Input : a graph $G = \langle V, E \rangle$

Output : an upper bound on the min-cut

```
1 Function Contract( $G, e$ ):  
2   | remove all edges between  $e.s$  and  $e.t$ ;  
3   | foreach edge  $e'$  containing  $e.s$  do  
4   |   | disconnect  $e'$  from  $e.s$  and reconnect it to  $e.t$ ;  
5   | end foreach  
6   |  $S_{e.t} \leftarrow S_{e.t} \cup S_{e.s}$ ; remove vertex  $e.s$  from  $V$ ;  
7   | return  $G$ ;  
8 end  
9 foreach  $v \in G.V$  do  $S_v = \{v\}$ ;  
10 while  $G$  has more than two vertices do  
11   |  $e \leftarrow \text{rand}(G.E)$ ;  $G \leftarrow \text{Contract}(G, e)$ ;           /* edge  $e = (e.s, e.t)$  */  
12 end while  
13 return  $|G.E|$ ;
```

Apply Karger's algorithm to the following graph G and find two different final values for $|G.E|$.



The process described in the function `Contract` of Karger's algorithm (5.21) is called *edge contraction*.

As observed in the previous exercise (5.22) running Karger's algorithm can lead to various values. The question is then to figure out how to retrieve the right one with high probability. In order to do this we first evaluate the cost of a contraction.

Lemma

Let $G = \langle V, E \rangle$ be a multigraph. A single contraction in Karger's algorithm takes $\mathcal{O}(|V|)$ time and the whole algorithm $\mathcal{O}(|V|^2)$.

Proof. For an edge $e = (s, t)$, a contraction reattaches all edges incident on s to t . In term of adjacency matrix we replace row t with the sum of rows s and t and column t with the sum of columns s and t . Then both row s and column s are cleared up. This can be performed in $\mathcal{O}(|V|)$ time, meaning Karger's algorithm has complexity $\mathcal{O}(|V|^2)$. \square

Lemma

Let $G = \langle V, E \rangle$ be a multigraph, and (S, T) be a min-cut. Then

$$\Pr [\text{Karger's algorithm ends with } (S, T)] \geq \frac{1}{\binom{|V|}{2}}.$$

Proof. First observe that if an edge (u, v) is contracted then only the cuts containing both u and v are unchanged. Therefore for Karger's algorithm to succeed the minimum cut (S, T) must remain untouched over all the random edge selections. Denoting the number of vertices by n , $n - 2$ contractions are to be performed in order to have only two vertices left. Naming those edges $\{e_1, \dots, e_{n-2}\}$, the goal is to determine the probability to choose a proper edge at each iteration of the algorithm.

Let k denote the size of a minimum cut in G . Then the minimum degree of any vertex in G is at least k , otherwise a cut of size less than k could be exhibited by disconnecting a vertex of degree less than k . Therefore G has at least $nk/2$ edges.

After a contraction the new graph has one less vertex but the degree of all the vertices remains at least k . So after i steps there are $n - i$ vertices and at least $(n - i)k/2$ edges.

Since any vertex has degree at least k the probability of selecting a “bad edge”, assuming none has been chosen before, is $2/(n - i)$.

Hence we can determine the probability of never choosing a “bad edge” during the whole process and thus end with (S, T) .

This probability is given by

$$\begin{aligned}\Pr [\text{find } (S, T)] &= \Pr [e_1, \dots, e_{n-2} \notin (S, T)] \\&= \Pr [e_1 \notin (S, T)] \prod_{i=1}^{n-3} \Pr [e_{i+1} \notin (S, T) \mid e_1, \dots, e_i \notin (S, T)] \\&\geq \prod_{i=0}^{n-3} \left(1 - \frac{2}{n-i}\right) \\&= \frac{2}{n(n-1)} = \frac{1}{\binom{n}{2}}.\end{aligned}$$



Since the probability of finding a min-cut is least $1/\binom{n}{2}$ it suffices to run the algorithm $l\binom{n}{2}$, for some value l . The probability of a run to succeed is then at least

$$1 - \left(1 - \frac{1}{\binom{n}{2}}\right)^{l\binom{n}{2}} \geq 1 - e^{-l}.$$

Therefore an appropriate choice for l is $c \ln n$, which leads to an error probability of at most $1/n^c$. Hence the total running time of Karger's algorithm is $\mathcal{O}(n^4 \log n)$.

Although this approach works well, it remains slow. The main reason is the random choice of the edge: at the beginning the multigraph features many edges and the probability of selecting an edge from a minimum cut is low. However as the process advances the probability of contracting an edge in the minimum cut grows.

Algorithm. (*Karger-Stein*)

Input : a graph $G = \langle V, E \rangle$

Output : a mini-cut in G

```
1 Function FastCut( $G$ ):  
2   if  $|G.V| > 6$  then  
3      $G_1 \leftarrow G$ ;  $G_2 \leftarrow G$ ;  
4      $t \leftarrow 1 + \frac{|G.V|}{\sqrt{2}}$ ;  
5     while  $|G_1.V| \geq t$  do  $e \leftarrow \text{rand}(G_1.E)$ ;  $G_1 \leftarrow \text{Contract}(G_1, e)$ ;  
6     while  $|G_2.V| \geq t$  do  $e \leftarrow \text{rand}(G_2.E)$ ;  $G_2 \leftarrow \text{Contract}(G_2, e)$ ;  
7     FastCut( $G_1$ ); FastCut( $G_2$ );  
8   else  
9     find the min-cut by enumeration  
10  end if  
11  return  $|G_1.E| \leq |G_2.E| ? |G_1.E| : |G_2.E|$   
12 end
```

Theorem

Given a multigraph with n vertices, Karger-Stein's algorithm discovers a minimum cut in time $\mathcal{O}(n^2 \log^3 n)$, with high probability.

Sketch of proof. First note that $6 \ll |V|$ and as such finding a minimum cut by enumeration only impacts the final complexity by a constant factor.

Given a cut, observe that the probability that it survives down to t vertices is at least $\binom{t}{2} / \binom{n}{2}$. Thus for $t = n/\sqrt{2}$ the probability of success is larger than $1/2$.

Since Karger-Stein's algorithm follows a divide and conquer strategy, its complexity can be expressed by a recurrence relation.

Then recalling that a single edge contraction costs $\mathcal{O}(n^2)$ (lemma 5.23) we get

$$T(n) = 2 \left(n^2 + T \left(\frac{n}{\sqrt{2}} \right) \right).$$

Hence by the master theorem (2.31) we conclude that the running time of Karger-Stein's algorithm is $\mathcal{O}(n^2 \log n)$.

We now consider the success probability. First as we start with n vertices and go down to $t = n/\sqrt{2}$, the success probability is $(t/n)^2 \approx 1/2$. Then at the next recursion level the graph shrinks from $n/\sqrt{2}$ to $n/2$ vertices, which means an overall success probability of about $1/4$.

More generally assume the minimum cut to still be in the graph and let $P(t)$ be the probability that a call to the algorithm with t vertices successfully computes it. Then G_i , $1 \leq i \leq 2$ still contains it with probability larger than a half. Therefore the probability that a recursive call succeeds is $1/2P(t/\sqrt{2})$. And since two recursive calls are performed

$$P(t) = 1 - \left(1 - \frac{1}{2}P\left(\frac{t}{\sqrt{2}}\right)\right)^2.$$

Solving this recurrence relation yields $P(n) = \Omega(1/\log n)$.¹ This means that Karger-Stein's algorithm needs to be run about $\log^2 n$ times in order to have an error probability of at most $\mathcal{O}(1/n)$ of preserving the minimum cut. This gives a final complexity of $\mathcal{O}(n^2 \log^3 n)$. \square

¹This result is proven in the homework.

Randomized algorithms:

- Bring much flexibility compared to deterministic ones
- Are often faster than deterministic ones
- Introduce imprecision on the output or on the complexity
- Require a good knowledge of probability theory
- Have proof that are often complex, even if the algorithm can be simply expressed

1001011111010100010101001111000111000100010111000000100110100001111101
1010010110011100111011101010001110000010001101110101011011111101010000
11011100111000110100110101100110110101010100111101000010101010101000
00011001111000100000011010101101111011110011000000101010101110100001
1100101110000100001001010110000011100010101000100110110010001101100100
0101011010111000001010110010010111000001010011110000001100111001110100
011000111110100000110100111110000110111001100110111101001101
1110100100000100111110111101001100001100110010010011111010001001001
1001111000111110100101011101001111101000111100111101110100010
01111101111110011111001001001001101001100100100001111110000110
11101101100111100001001010100110110011010111111011111010
1101101000100110010100100111100001010011100001000101110010011011010011
0000110101010001100110000111000100101110011110000101001101100100000110
1101111010010101011101011111000010100010011101100000010110111100000011
0111101000100101011011001110011101011011110010110101011001100110101100
10110110110111110100000000111011100011110101110110011100111010011100000

Thank you!