Clique

CLIQUE = \{ (G, k) \mid G \text{ is a graph with a clique of size } \geq k \}

(clique = set of vertices every pair of which are connected by an edge)

• CLIQUE is \textbf{NP}-complete.

Circuit lower bounds

• We think that \textbf{NP} requires exponential-size circuits.
• Where should we look for a problem to attempt to prove this?
• Intuition: “hardest problems” – i.e., \textbf{NP}-complete problems

Monotone problems

• Definition: monotone language = language \( L \subseteq \{0,1\}^* \)
  such that \( x \in L \) implies \( x' \in L \) for all \( x \preceq x' \).
  
  – flipping a bit of the input from 0 to 1 can only change the output from “no” to “yes” (or not at all)

• some \textbf{NP}-complete languages are monotone
  – e.g. CLIQUE (given as adjacency matrix):
  – others: HAMILTON CYCLE, SET COVER…
  – but not SAT, KNAPSACK…
Monotone circuits

A restricted class of circuits:

- Definition: monotone circuit = circuit whose gates are ANDs (\(\land\)), ORs (\(\lor\)), but no NOTs
- can compute exactly the monotone fn.
  - monotone functions closed under AND, OR

Monotone circuits

A question:

Do all poly-time computable monotone functions have poly-size monotone circuits?

- recall: true in non-monotone case

Monotone circuits

A monotone circuit for CLIQUE\(_{n,k}\):

- Input: graph \(G = (V,E)\) as adj. matrix, \(|V| = n\)
  - variable \(x_{i,j}\) for each possible edge \((i,j)\)
- ISCLIQUE(S) = monotone circuit that = 1 iff \(S \subseteq V\) is a clique: \(\land_{i,j \in S} x_{i,j}\)
- CLIQUE\(_{n,k}\) computed by monotone circuit:
  \(\lor_{S \subseteq V, |S| = k} ISCLIQUE(S)\)

Monotone circuits

Size of this monotone circuit for CLIQUE\(_{n,k}\):

- when \(k = n^{1/4}\), size is approximately:
  \(\left(\frac{n}{n^{1/4}}\right)^{n^{1/4}} \left(\frac{n^{1/4}}{2}\right)^2 \approx n^{1/4} n^{1/4}\)

Theorem (Razborov 85): monotone circuits for CLIQUE\(_{n,k}\) with \(k = n^{1/4}\) must have size at least

\[2^{\Omega(n^{1/8})}.\]

Proof idea

- "method of approximation"
- suppose \(C\) is a monotone circuit for CLIQUE\(_{n,k}\)
- build another monotone circuit \(CC\) that "approximates" \(C\) gate-by-gate
Proof idea

- on test collection of positive/negative instances of $\text{CLIQUE}_{n,k}$:
  - local property: few errors at each gate
  - global property: many errors on test collection

- Conclude: C has many gates

Notation

- input: graph $G = (V, E)$
- variable $x_{j,k}$ for each potential edge $(j, k)$
- $\text{CC}(X_1, X_2, \ldots X_m)$, where $X_i \subseteq V$, means:
  $\forall i \left( \bigwedge_{j,k \in X_i} x_{j,k} \right)$

  - For example: $\text{CC}(X_1, X_2, \ldots X_m)$ where the $X_i$ range over all $k$-subsets of $V$
  - this is the obvious monotone circuit for $\text{CLIQUE}_{n,k}$ from a previous slide.

  $\left[ \text{CC}() = 0; \left( \bigwedge_{j,k \in \emptyset} x_{j,k} \right) = 1 \right]$

Preview

- approximate circuit $\text{CC}(X_1, X_2, \ldots X_m)$
  - $n =$ # nodes
  - $k = n^{1/4} =$ size of clique
  - $h = n^{1/8} =$ max. size of subsets $X_i$
    - this is “global property” that ensures lots of errors
    - many graphs $G$ with no $k$-cliques, but clique on $X_i$ of size $h$

Building CC

- $\text{CC}$ (“crude circuit”) for circuit $C$ defined inductively as follows:
  - $\text{CC}$ for single variable $x_{j,k}$ is just $\text{CC}(\{ j, k \})$
    - no errors yet!
  - $\text{CC}$ for circuit $C$ of form:
    - “approximate OR” of $\text{CC}$ for $C'$, $\text{CC}$ for $C''$

- $\text{CC}$ for circuit $C$ of form:
  - “approximate AND” of $\text{CC}$ for $C'$, $\text{CC}$ for $C''$
  - “approximate OR” and “approximate AND” steps introduce errors
Approximate OR

\[ \text{CC}(X_1, X_2, \ldots, X_m) \]  
\[ \text{CC}(Y_1, Y_2, \ldots, Y_m) \]

- **exact OR:**
  \[ \text{CC}(X_1, X_2, \ldots, X_m, Y_1, Y_2, \ldots, Y_m) \]
  - set sizes still \( \leq h \)
  - may be up to \( 2M \) sets; need to reduce to \( M \)

- **approximate OR:**
  \[ \text{CC}(\text{pluck}(X_1, X_2, \ldots, X_m, Y_1, Y_2, \ldots, Y_m)) \]

Approximate OR

- throw away sets? **bad:** many errors
- throw away overlapping sets? – **better**

- throw away special configuration of overlapping sets – best

Sunflowers

- **Definition:** \((h, p)-\text{sunflower}\) is a family of \( p \) sets ("petals") each of size at most \( h \), such that intersection of every pair is a subset \( S \) (the "core").

**Lemma** (Erdős-Rado): Every family of more than \( M = (p-1)^{h}h! \) sets, each of size at most \( h \), contains an \((h, p)-\text{sunflower}\).

- **Proof:**
  - not hard
  - in Papadimitriou, elsewhere

Approximate AND

\[ \text{CC}(X_1, X_2, \ldots, X_m) \]  
\[ \text{CC}(Y_1, Y_2, \ldots, Y_m) \]

- **exact AND:**
  \[ \text{CC}(\{ (X_i, Y_j) : 1 \leq i \leq m', 1 \leq j \leq m'' \}) \]
  - some sets may be larger than \( h \); discard them
  - may be up to \( M^2 \) sets. While > \( M \) sets, find \((h, p)-\text{sunflower}\); replace with its core ("pluck")

- **approximate AND:**
  \[ \text{CC}(\text{pluck}(\{ (X_i, Y_j) : |X_i \cup Y_j| \leq h \})) \]
Test collection

- **Positive instances**: all graphs $G$ on $n$ nodes with a $k$-clique and no other edges.

| $G$ |

Test collection

- **Negative instances**: $(k\text{-})$-partite graph
  - $k$ colors
  - color each node uniformly at random with one of the colors
  - edge $(x, y)$ iff $x, y$ different colors
  - no $k$-clique
  - include graphs in their multiplicities
  - makes analysis easier

"Local" analysis

- "false positive":
  - negative example
  - gate is supposed to output 0, but our CC outputs 1

**Lemma**: each approximation step introduces at most $M^2(k\text{-}1)^n/2^p$ false positives.

"Local" analysis

- Proof:
  - case 1: OR
  - $CC(X_1, X_2, \ldots, X_m)$
  - $CC(Y_1, Y_2, \ldots, Y_m)$
  - $CC(pluck(X_1, X_2, \ldots, X_m, Y_1, Y_2, \ldots, Y_m))$

  - given "plucking": replace $Z_1 \ldots Z_p$ with $Z$
  - bad case: clique on $Z$, and each petal is missing at least one edge

"Local" analysis

- what is the probability of a repeated color in each $Z_i$ but no repeated colors in $Z$?
  - $Pr[R(Z_1) \land R(Z_2) \ldots R(Z_p) \land \neg R(Z)]$
  - $\leq Pr[R(Z_1) \land R(Z_2) \ldots R(Z_p) \land \neg R(Z)]$
  - (definition of conditional probability)
  - $= \prod_i Pr[R(Z_i) | \neg R(Z)]$
  - (independent events given no repeats in $Z$)
  - $\leq \prod_i Pr[R(Z_i)]$
  - (obviously larger)

"Local" analysis

- for every pair of vertices in $Z_i$, probability of same color is $1/(k\text{-}1)$
  - $R(Z_i) \leq (h \text{ choose } 2)/(k\text{-}1) \leq \frac{1}{2}$
  - $\prod_i Pr[R(Z_i)] \leq \left(\frac{1}{2}\right)^p$
  - # negative examples is $(k\text{-}1)^n$
  - # false positives in given plucking step is at most $(\frac{1}{2})^p(k\text{-}1)^n$
  - at most $M$ plucking steps
  - # false positives at OR $\leq M\left(\frac{1}{2}\right)^p(k\text{-}1)^n$
“Local” analysis

– case 2: AND

\[
\text{CC}(X_1, X_2, \ldots, X_m) \quad \text{CC}(Y_1, Y_2, \ldots, Y_m) \\
\text{CC}(\text{pluck}(\{(X_i \cup Y_j) : |X_i \cup Y_j| \leq h\}))
\]

– discarding sets \((X_i \cup Y_j)\) larger than \(h\) can only make circuit accept fewer examples
  • no false positives here

“Local” analysis

– up to \(M^2\) pluckings
  – each introduces at most \(\left(\frac{1}{2}\right)^p(k-1)^n\) false positives (previous slides)
  – # false positives at AND \(\leq M^2\left(\frac{1}{2}\right)^p(k-1)^n\)

“Local” analysis

• “false negative”:
  – positive example;
  – gate is supposed to output 1, but our CC outputs 0

Lemma: each approximation step introduces at most \(M\left(\frac{n-h-1}{k-h-1}\right)\) false negatives.

“Local” analysis

– discarding set \(Z = (X_i \cup Y_j)\) larger than \(h\) may introduce false negatives
  – any clique that includes \(Z\) is a problem; there are at most \(\frac{n-|Z|}{k-|Z|} \leq \frac{n-h-1}{k-h-1}\)
  such positive examples, since \(|Z|>h\)
  – at most \(M^2\) such deletions
  – we’ve seen plucking doesn’t matter

“Global” analysis

Lemma: every non-trivial CC outputs 1 on at least \(\frac{1}{2}\) of the negative examples.

• Proof:
  – CC contains some set \(X\) of size at most \(h\)
  – accepts all neg. examples with different colors in \(X\)
  – probability \(X\) has repeated colors is \(R(X) \leq \binom{h}{2}/(k-1) \leq \frac{1}{2}\)
  – probability over negative examples that CC accepts is at least \(\frac{1}{2}\).
Finishing up

• First possibility: trivial CC, rejects all positive examples
  – every positive example must have been false negative at some gate
  – number of gates must be at least:

\[
\binom{n}{k}/M^2 \binom{n-h-1}{k-h-1}
\]

\[
\frac{1}{2}(k-1)^2 / M^2 2^{-p(k-1)^2}
\]

Both quantities are at least \(2^{\Omega(n^{1/8})}\)

Finishing up

• Second possibility: CC accepts at least \(1/2\) of negative examples
  – every negative example must have been false positive at some gate
  – number of gates must be at least:

\[
\frac{1}{2}(k-1)^2 / M^2 2^{-p(k-1)^2}
\]

Conclusions

• A question (true in non-monotone case):
  Do all poly-time computable monotone functions have poly-size monotone circuits?
  • if yes, then we would have just proved \(P \neq NP\)
    – why?

• unfortunately, answer is no

• Razborov later showed similar (super-polynomial) lower bound for MATCHING, which is in \(P\)...
1. Communication complexity

- **Goal**: compute \( f(x, y) \) while communicating as few bits as possible between Alice and Bob
- count number of bits exchanged (computation free)
- at each step: one party sends bits that are a function of held input and received bits so far

\[
\text{two parties: Alice and Bob} \\
\text{function } f : \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\} \\
\text{Alice holds } x \in \{0,1\}^n; \text{ Bob holds } y \in \{0,1\}^n
\]

- simple function (equality):
  \( \text{EQ}(x, y) = 1 \) iff \( x = y \)
- simple protocol:
  – Alice sends \( x \) to Bob (\( n \) bits)
  – Bob sends \( \text{EQ}(x, y) \) to Alice (1 bit)
  – total: \( n + 1 \) bits
  – (works for any predicate \( f \))

Communication complexity

- Can we do better?
  – deterministic protocol?
  – probabilistic protocol?
  \( \text{at each step: one party sends bits that are a function of held input and received bits so far and the result of some coin tosses} \)
  \( \text{required to output } f(x, y) \) with high probability over all coin tosses

**Theorem**: no deterministic protocol can compute \( \text{EQ}(x, y) \) while exchanging fewer than \( n + 1 \) bits.

- Proof:
  – “input matrix”:
    \[
    X = \{0,1\}^n \\
    Y = \{0,1\}^n \\
    f(x,y)
    \]
  – assume 1 bit sent at a time, alternating (same proof works in general setting)
  – A sends 1 bit depending only on \( x \):
    \[
    X = \{0,1\}^n \\
    Y = \{0,1\}^n \\
    \text{inputs } x \text{ causing } A \text{ to send 1} \\
    \text{inputs } x \text{ causing } A \text{ to send 0}
    \]

- B sends 1 bit depending only on \( y \) and received bit:
  \[
  X = \{0,1\}^n \\
  Y = \{0,1\}^n \\
  \text{inputs } y \text{ causing } B \text{ to send 1} \\
  \text{inputs } y \text{ causing } B \text{ to send 0}
  \]
Communication complexity

– at end of protocol involving $k$ bits of communication, matrix is partitioned into at most $2^k$ combinatorial rectangles

– bits sent in protocol are the same for every input $(x, y)$ in given rectangle
– conclude: $f(x, y)$ must be constant on each rectangle

Communication complexity

– any partition into combinatorial rectangles with constant $f(x, y)$ must have $2^n+1$ rectangles
– protocol that exchanges $\leq n$ bits can only create $2^n$ rectangles, so must exchange at least $n+1$ bits.

Communication complexity

– protocol for EQ employing randomness?
– Alice picks random prime $p$ in $\{1...4n^2\}$, sends:
  * $p$
  * $(x \mod p)$
– Bob sends:
  * $(y \mod p)$
– players output 1 if and only if:
  $(x \mod p) = (y \mod p)$

Communication complexity

– $O(\log n)$ bits exchanged
– if $x = y$, always correct
– if $x \neq y$, incorrect if and only if:
  $p$ divides $|x - y|
– # primes in range is $\geq 2n$
– # primes dividing $|x - y|$ is $\leq n$
– probability incorrect $\leq 1/2$

Randomness gives an exponential advantage!!

2. Polynomial identity testing

• Given: polynomial $p(x_1, x_2, \ldots, x_n)$ as arithmetic formula (fan-out 1):
  * multiplication (fan-in 2)
  * addition (fan-in 2)
  * negation (fan-in 1)

• Question: Is $p$ identically zero?
  – i.e., is $p(x) = 0$ for all $x \in F^n$
  – (assume $F$ larger than degree…)

“polynomial identity testing” because given two polynomials $p, q$, we can check the identity $p = q$ by checking if $(p - q) = 0$
Polynomial identity testing

• try all \(|F|^n\) inputs?
  – may be exponentially many
• multiply out symbolically, check that all coefficients are zero?
  – may be exponentially many coefficients
• can randomness help?
  – i.e., flip coins, allow small probability of wrong answer

Lemma (Schwartz-Zippel): Let 
  \[ p(x_1, x_2, \ldots, x_n) \]
be a total degree \(d\) polynomial over a field \(F\) and let \(S\) be any subset of \(F\). Then if \(p\) is not identically 0,
  \[ \Pr_{r_1, r_2, \ldots, r_n \in S} [p(r_1, r_2, \ldots, r_n) = 0] \leq d/|S|. \]