Sunflowers

• Fix $p \in \mathbb{Z}^+$ and $\ell \in \mathbb{Z}^+$.

• A sunflower is a family of $p$ sets $\{ P_1, P_2, \ldots, P_p \}$, called petals, each of cardinality at most $\ell$.

• Furthermore, all pairs of sets in the family must have the same intersection (called the core of the sunflower).
A Sample Sunflower

\[ \{\{1, 2, 3, 5\}, \{1, 2, 6, 9\}, \{0, 1, 2, 11\}, \{1, 2, 12, 13\}, \{1, 2, 8, 10\}, \{1, 2, 4, 7\}\}. \]
The Erdős-Rado Lemma

**Lemma 88** Let $\mathcal{Z}$ be a family of more than $M \triangleq (p - 1)^\ell \ell!$ nonempty sets, each of cardinality $\ell$ or less. Then $\mathcal{Z}$ must contain a sunflower (with $p$ petals).

- Induction on $\ell$.
- For $\ell = 1$, $p$ different singletons form a sunflower (with an empty core).
- Suppose $\ell > 1$.
- Consider a maximal subset $\mathcal{D} \subseteq \mathcal{Z}$ of disjoint sets.
  - Every set in $\mathcal{Z} - \mathcal{D}$ intersects some set in $\mathcal{D}$.
The Proof of the Erdős-Rado Lemma (continued)

For example,

\[\mathcal{Z} = \{\{1, 2, 3, 5\}, \{1, 3, 6, 9\}, \{0, 4, 8, 11\}, \{4, 5, 6, 7\}, \{5, 8, 9, 10\}, \{6, 7, 9, 11\}\},\]

\[\mathcal{D} = \{\{1, 2, 3, 5\}, \{0, 4, 8, 11\}\}.\]
The Proof of the Erdős-Rado Lemma (continued)

• Suppose $\mathcal{D}$ contains at least $p$ sets.
  
  $\quad$ $\mathcal{D}$ constitutes a sunflower with an empty core.

• Suppose $\mathcal{D}$ contains fewer than $p$ sets.
  
  $\quad$ Let $\mathcal{C}$ be the union of all sets in $\mathcal{D}$.
  $\quad$ $|\mathcal{C}| \leq (p - 1)\ell$.
  $\quad$ $\mathcal{C}$ intersects every set in $\mathcal{Z}$ by $\mathcal{D}$’s maximality.
  $\quad$ There is a $d \in \mathcal{C}$ that intersects more than \[
  \frac{M}{(p-1)\ell} = (p - 1)^{\ell-1}(\ell - 1)! \text{ sets in } \mathcal{Z}.
  \]
  $\quad$ Consider $\mathcal{Z}' = \{ Z - \{ d \} : Z \in \mathcal{Z}, d \in Z \}$. 

The Proof of the Erdős-Rado Lemma (concluded)

• (continued)
  - \( \mathcal{Z}' \) has more than \( M' \triangleq (p - 1)^{\ell - 1}(\ell - 1)! \) sets.
  - \( M' \) is just \( M \) with \( \ell \) replaced with \( \ell - 1 \).
  - \( \mathcal{Z}' \) contains a sunflower by induction, say
    \[ \{ P_1, P_2, \ldots, P_p \} \]
  - Now,
    \[ \{ P_1 \cup \{ d \}, P_2 \cup \{ d \}, \ldots, P_p \cup \{ d \} \} \]
    is a sunflower in \( \mathcal{Z} \).
Comments on the Erdős-Rado Lemma

• A family of more than \( M \) sets must contain a sunflower.

• **Plucking** a sunflower means replacing the sets in the sunflower by its core.

• By *repeatedly* finding a sunflower and plucking it, we can reduce a family with more than \( M \) sets to a family with at most \( M \) sets.

• If \( \mathcal{Z} \) is a family of sets, the above result is denoted by \( \text{pluck}(\mathcal{Z}) \).

• \( \text{pluck}(\mathcal{Z}) \) is not unique.\(^a\)

\(^a\)It depends on the sequence of sunflowers one plucks. Fortunately, this issue is not material to the proof.
An Example of Plucking

- Recall the sunflower on p. 811:

\[ Z = \begin{array}{c}
\{ 1, 2, 3, 5 \}, \{ 1, 2, 6, 9 \}, \{ 0, 1, 2, 11 \}, \\
\{ 1, 2, 12, 13 \}, \{ 1, 2, 8, 10 \}, \{ 1, 2, 4, 7 \}\end{array} \]

- Then

\[ \text{pluck}(Z) = \{ \{ 1, 2 \} \}. \]
Razborov’s Theorem

Theorem 89 (Razborov, 1985) There is a constant $c$ such that for large enough $n$, all monotone circuits for $\text{CLIQUE}_{n,k}$ with $k = n^{1/4}$ have size at least $n^{cn^{1/8}}$.

- We shall approximate any monotone circuit for $\text{CLIQUE}_{n,k}$ by a restricted kind of crude circuit.
- The approximation will proceed in steps: one step for each gate of the monotone circuit.
- Each step introduces few errors (false positives and false negatives).
- Yet, the final crude circuit has exponentially many errors.
The Proof

- Fix $k = n^{1/4}$.
- Fix $\ell = n^{1/8}$.
- Note that

$$2 \binom{\ell}{2} \leq k - 1.$$

- $p$ will be fixed later to be $n^{1/8} \log n$.
- Fix $M = (p - 1)^{\ell} \ell!$.
  - Recall the Erdős-Rado lemma (p. 812).

\[ \text{Corrected by Mr. Moustapha Bande (D98922042) on January 5, 2010.} \]
The Proof (continued)

- Each crude circuit used in the approximation process is of the form \( CC(X_1, X_2, \ldots, X_m) \), where:
  - \( X_i \subseteq V \).
  - \( |X_i| \leq \ell \).
  - \( m \leq M \).

- It answers true if any \( X_i \) is a clique.

- We shall show how to approximate any monotone circuit for \( \text{CLIQUE}_{n,k} \) by such a crude circuit, inductively.

- The induction basis is straightforward:
  - Input gate \( g_{ij} \) is the crude circuit \( CC(\{i, j\}) \).
The Proof (continued)

• A monotone circuit is the OR or AND of two subcircuits.

• We will build approximators of the overall circuit from the approximators of the two subcircuits.
  – Start with two crude circuits CC(\mathcal{X}) and CC(\mathcal{Y}).
  – \mathcal{X} and \mathcal{Y} are two families of at most M sets of nodes, each set containing at most \ell nodes.
  – We will construct the approximate OR and the approximate AND of these subcircuits.
  – Then show both approximations introduce few errors.
The Proof: OR

- **CC(\(\mathcal{X} \cup \mathcal{Y}\))** is *equivalent to* the OR of **CC(\(\mathcal{X}\))** and **CC(\(\mathcal{Y}\)).**
  - Trivially, a node set \(C \in \mathcal{X} \cup \mathcal{Y}\) is a clique if and only if \(C \in \mathcal{X}\) is a clique or \(C \in \mathcal{Y}\) is a clique.

- Violations in using **CC(\(\mathcal{X} \cup \mathcal{Y}\))** occur when \(|\mathcal{X} \cup \mathcal{Y}| > M\).

- Such violations are eliminated by using

\[
CC(\text{pluck}(\mathcal{X} \cup \mathcal{Y}))
\]

as the approximate OR of **CC(\(\mathcal{X}\))** and **CC(\(\mathcal{Y}\)).**
The Proof: OR (continued)

- If $CC(\mathcal{Z})$ is true, then $CC(\text{pluck}(\mathcal{Z}))$ must be true.
  - The quick reason: If $Y$ is a clique, then a subset of $Y$ must also be a clique.
  - Let $Y \in \mathcal{Z}$ be a clique.
  - There must exist an $X \in \text{pluck}(\mathcal{Z})$ such that $X \subseteq Y$.
  - This $X$ is also a clique.
The Proof: $OR$ (continued)
The Proof: $\text{OR}$ (concluded)

- $\text{CC}(\text{pluck}(\mathcal{X} \cup \mathcal{Y}))$ introduces a **false positive** if a negative example makes both $\text{CC}(\mathcal{X})$ and $\text{CC}(\mathcal{Y})$ return false but makes $\text{CC}(\text{pluck}(\mathcal{X} \cup \mathcal{Y}))$ return true.

- $\text{CC}(\text{pluck}(\mathcal{X} \cup \mathcal{Y}))$ introduces a **false negative** if a positive example makes either $\text{CC}(\mathcal{X})$ or $\text{CC}(\mathcal{Y})$ return true but makes $\text{CC}(\text{pluck}(\mathcal{X} \cup \mathcal{Y}))$ return false.

- We next count the number of false positives and false negatives introduced\(^a\) by $\text{CC}(\text{pluck}(\mathcal{X} \cup \mathcal{Y}))$.

- Let us work on false negatives for $\text{OR}$ first.

---

\(^a\)Compared with $\text{CC}(\mathcal{X} \cup \mathcal{Y})$ of course.
The Number of False Negatives\textsuperscript{a}

Lemma 90 \( CC(\text{pluck}(\mathcal{X} \cup \mathcal{Y})) \) introduces no false negatives.

- Each plucking replaces sets in a crude circuit by their common subset.

- This makes the test for cliqueness less stringent.\textsuperscript{b}

\textsuperscript{a}CC(\text{pluck}(\mathcal{X} \cup \mathcal{Y})) introduces a false negative if a positive example makes either \( CC(\mathcal{X}) \) or \( CC(\mathcal{Y}) \) return true but makes \( CC(\text{pluck}(\mathcal{X} \cup \mathcal{Y})) \) return false.

\textsuperscript{b}Recall p. 823.
The Number of False Positives

Lemma 91 \( CC(\text{pluck}(\mathcal{X} \cup \mathcal{Y})) \) introduces at most 
\[
\frac{2^M}{p-1} 2^{-p}(k - 1)^n \text{ false positives.}
\]

- Each plucking operation replaces the sunflower \( \{ Z_1, Z_2, \ldots, Z_p \} \) with its common core \( Z \).

- A false positive is \textit{necessarily} a coloring such that:
  - There is a pair of identically colored nodes in \textit{each} petal \( Z_i \) (and so \( CC(Z_1, Z_2, \ldots, Z_p) \) returns false).
  - But the core contains distinctly colored nodes (thus forming a clique).
  - This implies at least one node from each identical-color pair was plucked away.
Proof of Lemma 91 (continued)
Proof of Lemma 91 (continued)

• We now count the number of such colorings.
• Color nodes in $V$ at random with $k - 1$ colors.
• Let $R(X)$ denote the event that there are repeated colors in set $X$. 
Proof of Lemma 91 (continued)

• Now

\[
\begin{align*}
\text{prob}[ R(Z_1) \land \cdots \land R(Z_p) \land \neg R(Z) ] & \quad (24) \\
\leq & \quad \text{prob}[ R(Z_1) \land \cdots \land R(Z_p) | \neg R(Z) ] \\
= & \quad \prod_{i=1}^{p} \text{prob}[ R(Z_i) | \neg R(Z) ] \\
\leq & \quad \prod_{i=1}^{p} \text{prob}[ R(Z_i) ]. \quad (25)
\end{align*}
\]

– First equality holds because \( R(Z_i) \) are independent given \( \neg R(Z) \) as \( Z \) contains their only common nodes.

– Last inequality holds as the likelihood of repetitions in \( Z_i \) decreases given no repetitions in its subset \( Z \).
Proof of Lemma 91 (continued)

- Consider two nodes in $Z_i$.
- The probability that they have identical color is
  \[ \frac{1}{k - 1}. \]
- Now
  \[ \text{prob}[R(Z_i)] \leq \frac{|Z_i|}{k - 1} \leq \frac{\ell}{k - 1} \leq \frac{1}{2}. \]
- So the probability\(^a\) that a random coloring is a new false positive is at most $2^{-p}$ by inequality (25) on p. 830.

\(^a\)Proportion, if you so prefer.
Proof of Lemma 91 (continued)

• As there are \((k - 1)^n\) different colorings, each plucking introduces at most \(2^{-p}(k - 1)^n\) false positives.

• Recall that \(|\mathcal{X} \cup \mathcal{Y}| \leq 2M\).

• When the procedure \(\text{pluck}(\mathcal{X} \cup \mathcal{Y})\) ends, the set system contains \(\leq M\) sets.
Proof of Lemma 91 (concluded)

- Each plucking reduces the number of sets by $p - 1$.
- Hence at most $2M/(p - 1)$ pluckings occur in $\text{pluck}(\mathcal{X} \cup \mathcal{Y})$.
- At most
  \[
  \frac{2M}{p - 1} 2^{-p(k - 1)n}
  \]
  false positives are introduced.\(^a\)

---

\(^a\)Note that the numbers of errors are added not multiplied. Recall that we count how many new errors are introduced by each approximation step. Contributed by Mr. Ren-Shuo Liu (D98922016) on January 5, 2010.
The Proof: AND

- The approximate AND of crude circuits $CC(\mathcal{X})$ and $CC(\mathcal{Y})$ is
  
  $$CC(\text{pluck}(\{ X_i \cup Y_j : X_i \in \mathcal{X}, Y_j \in \mathcal{Y}, |X_i \cup Y_j| \leq \ell \})).$$

- We need to count the number of errors this approximate AND introduces on the positive and negative examples.
The Proof: AND (continued)

- The approximate AND introduces a **false positive** if a negative example makes either $\text{CC}(\mathcal{X})$ or $\text{CC}(\mathcal{Y})$ return false but makes the approximate AND return true.

- The approximate AND introduces a **false negative** if a positive example makes both $\text{CC}(\mathcal{X})$ and $\text{CC}(\mathcal{Y})$ return true but makes the approximate AND return false.

- Introduction of errors means we ignore scenarios where the AND of $\text{CC}(\mathcal{X})$ and $\text{CC}(\mathcal{Y})$ is already wrong.
The Proof: AND (continued)

- $\text{CC}(\{ X_i \cup Y_j : X_i \in \mathcal{X}, Y_j \in \mathcal{Y} \})$ introduces no false positives and no false negatives over our positive and negative examples.\(^a\)
  - Suppose $\text{CC}(\{ X_i \cup Y_j : X_i \in \mathcal{X}, Y_j \in \mathcal{Y} \})$ returns true.
  - Then some $X_i \cup Y_j$ is a clique.
  - Thus $X_i \in \mathcal{X}$ and $Y_j \in \mathcal{Y}$ are cliques, making both $\text{CC}(\mathcal{X})$ and $\text{CC}(\mathcal{Y})$ return true.
  - So $\text{CC}(\{ X_i \cup Y_j : X_i \in \mathcal{X}, Y_j \in \mathcal{Y} \})$ introduces no false positives.

\(^a\)Unlike the OR case on p. 822, we are not claiming that $\text{CC}(\{ X_i \cup Y_j : X_i \in \mathcal{X}, Y_j \in \mathcal{Y} \})$ is equivalent to the AND of $\text{CC}(\mathcal{X})$ and $\text{CC}(\mathcal{Y})$. Equivalence is more than we need in either case.
The Proof: AND (concluded)

• (continued)
  – On the other hand, suppose both CC(\mathcal{X}) and CC(\mathcal{Y}) accept a positive example with a clique \mathcal{C} of size \( k \).
  – This clique \mathcal{C} must contain an \( X_i \in \mathcal{X} \) and a \( Y_j \in \mathcal{Y} \).
  – As this clique \mathcal{C} also contains \( X_i \cup Y_j \),\(^a\) the new circuit returns true.
  – CC(\{ \( X_i \cup Y_j \) : X_i \in \mathcal{X}, Y_j \in \mathcal{Y} \}) introduces no false negatives.

• We now bound the number of false positives and false negatives introduced\(^b\) by the approximate AND.

\(^a\)See next page.
\(^b\)Compared with CC(\{ \( X_i \cup Y_j \) : X_i \in \mathcal{X}, Y_j \in \mathcal{Y} \}) of course.
Clique of size $k$
The Number of False Positives

**Lemma 92** The approximate AND introduces at most $M^22^{-p}(k - 1)^n$ false positives.

- We prove this claim in stages.
- We already knew $\text{CC}({X_i \cup Y_j : X_i \in \mathcal{X}, Y_j \in \mathcal{Y}})$ introduces no false positives.\(^a\)
- $\text{CC}({X_i \cup Y_j : X_i \in \mathcal{X}, Y_j \in \mathcal{Y}, |X_i \cup Y_j| \leq \ell})$ introduces no additional false positives because we are testing potentially fewer sets for cliqueness.

\(^a\)Recall p. 836.
Proof of Lemma 92 (concluded)

- \(|\{ X_i \cup Y_j : X_i \in \mathcal{X}, Y_j \in \mathcal{Y}, |X_i \cup Y_j| \leq \ell \}| \leq M^2\).
- Each plucking reduces the number of sets by \(p - 1\).
- So pluck(\(\{ X_i \cup Y_j : X_i \in \mathcal{X}, Y_j \in \mathcal{Y}, |X_i \cup Y_j| \leq \ell \}\)) involves \(\leq M^2/(p - 1)\) pluckings.
- Each plucking introduces at most \(2^{-p}(k - 1)^n\) false positives by the proof of Lemma 91 (p. 827).
- The desired upper bound is

\[
\left[ \frac{M^2}{(p - 1)} \right] 2^{-p}(k - 1)^n \leq M^2 2^{-p}(k - 1)^n.
\]
The Number of False Negatives

Lemma 93 The approximate AND introduces at most $M^2 \binom{n-\ell-1}{k-\ell-1}$ false negatives.

- We again prove this claim in stages.
- We knew $\text{CC}\left(\{X_i \cup Y_j : X_i \in \mathcal{X}, Y_j \in \mathcal{Y}\}\right)$ introduces no false negatives.a

---

aRecall p. 836.
Proof of Lemma 93 (continued)

• \(\text{CC}\left(\{ X_i \cup Y_j : X_i \in \mathcal{X}, Y_j \in \mathcal{Y}, |X_i \cup Y_j| \leq \ell \}\right)\) introduces \(\leq M^2\binom{n-\ell-1}{k-\ell-1}\) false negatives.

  – Deletion of set \(Z \triangleq X_i \cup Y_j\) larger than \(\ell\) introduces false negatives only if \(Z\) is part of a clique.
  
  – There are \(\binom{n-|Z|}{k-|Z|}\) such cliques.
    
    * It is the number of positive examples whose clique contains \(Z\).

  – \(\binom{n-|Z|}{k-|Z|}\) \(\leq \binom{n-\ell-1}{k-\ell-1}\) as \(|Z| > \ell\).

  – There are at most \(M^2\) such \(Z\)s.
Proof of Lemma 93 (concluded)

- Plucking introduces no false negatives.
  - Recall that if $CC(\mathcal{Z})$ is true, then $CC(pluck(\mathcal{Z}))$ must be true.\(^a\)

\(^a\)Recall p. 823.
Two Summarizing Lemmas

From Lemmas 91 (p. 827) and 92 (p. 839), we have:

Lemma 94 Each approximation step introduces at most $M^22^{-p}(k - 1)^n$ false positives.

From Lemmas 90 (p. 826) and 93 (p. 841), we have:

Lemma 95 Each approximation step introduces at most $M^2\binom{n-\ell-1}{k-\ell-1}$ false negatives.
The Proof (continued)

- The above two lemmas show that each approximation step introduces “few” false positives and false negatives.

- We next show that the resulting crude circuit has “a lot” of false positives or false negatives.
The Final Crude Circuit

Lemma 96 Every final crude circuit is:

1. Identically false—thus wrong on all positive examples.
2. Or outputs true on at least half of the negative examples.

• Suppose it is not identically false.

• By construction, it accepts at least those graphs that have a clique on some set $X$ of nodes, with

$$|X| \leq \ell = n^{1/8} < n^{1/4} = k.$$
Proof of Lemma 96 (concluded)

• The proof of Lemma 91 (p. 827ff) shows that at least half of the colorings assign different colors to nodes in $X$.

• So at least half of the colorings — thus negative examples — have a clique in $X$ and are accepted.
The Proof (continued)

- Recall the constants on p. 819:

\[
\begin{align*}
  k & \triangleq n^{1/4}, \\
  \ell & \triangleq n^{1/8}, \\
  p & \triangleq n^{1/8} \log n, \\
  M & \triangleq (p - 1)^\ell \ell! < n^{(1/3)n^{1/8}} \text{ for large } n.
\end{align*}
\]
The Proof (continued)

- Suppose the final crude circuit is identically false.
  - By Lemma 95 (p. 844), each approximation step introduces at most $M^2\binom{n-\ell-1}{k-\ell-1}$ false negatives.
  - There are $\binom{n}{k}$ positive examples.
  - The original monotone circuit for $\text{CLIQUE}_{n,k}$ has at least
    \[
    \frac{\binom{n}{k}}{M^2\binom{n-\ell-1}{k-\ell-1}} \geq \frac{1}{M^2} \left(\frac{n-\ell}{k}\right)^\ell \geq n^{(1/12)n^{1/8}}
    \]
    gates for large $n$. 
The Proof (concluded)

• Suppose the final crude circuit is not identically false.
  – Lemma 96 (p. 846) says that there are at least \((k - 1)^n/2\) false positives.
  – By Lemma 94 (p. 844), each approximation step introduces at most \(M^22^{-p}(k - 1)^n\) false positives.
  – The original monotone circuit for \(\text{CLIQUE}_{n,k}\) has at least

\[
\frac{(k - 1)^n/2}{M^22^{-p}(k - 1)^n} = \frac{2^{p-1}}{M^2} \geq \frac{n^{(1/3)n^{1/8}}}{n^{(1/3)n^{1/8}}}
\]

gates.
Alexander Razborov (1963–)
P \neq NP Proved?

- Razborov’s theorem says that there is a monotone language in NP that has no polynomial monotone circuits.

- If we can prove that all monotone languages in P have polynomial monotone circuits, then P \neq NP.

- But Razborov proved in 1985 that some monotone languages in P have no polynomial monotone circuits!