The Probabilistic Method

Andreas Galanis

April 30, 2009

Table of contents

The Probabilistic Method

- Szele was the first to use it 1943.
- Paul Erdős first used it in 1947.
  - He used the method to its full extent, revealing the power of the method.

The Basic Method

1. Goal: Prove that a structure with certain desired properties exists.
2. Define an appropriate probability space of structures.
3. Show that the desired properties hold in this space with positive probability.

Useful facts:

$$\Pr\left[\bigcup A_i\right] \leq \sum \Pr[A_i]$$

Ramsey numbers

- $R(k, l)$ the smallest integer n such that in any two-coloring of the edges of a K_n by red and blue, either there is a red K_k or there is a blue K_l.
- $R(3, 3) = 6$.
- Determining the exact value $R(k, l)$ is very difficult.
- $43 \leq R(5, 5) \leq 49$ and that is the best we know.
- *Imagine an alien force, vastly more powerful than us landing on Earth and demanding the value of* $R(5, 5)$ *or they will destroy our planet. In that case, we should marshal all our computers and all our mathematicians and attempt to find the value. But suppose, instead, that they asked for* $R(6, 6)$*, we should attempt to destroy the aliens.*

A lower bound on $R(k, k)$

Claim: If $\binom{n}{k}2^{1-\binom{k}{2}} < 1$ then $R(k, k) > n$.

- Consider a random two-coloring of K_n, each edge colored independently red or blue.
- Consider any fixed set R of k vertices.
    - Let A_R denote the event that the graph induced by these k vertices is monochromatic.
- $\Pr[A_R] = 2^{1-\binom{k}{2}}$.
- $$\Pr\left[\bigcup_R A_R\right] \leq \sum_R \Pr[A_R] = \binom{n}{k}2^{1-\binom{k}{2}} < 1$$
- The probability that no A_R occurs is positive, i.e. there exists a coloring without a monochromatic K_k.

$R(k,k) > 2^{\left\lfloor \frac{k}{2} \right\rfloor}$

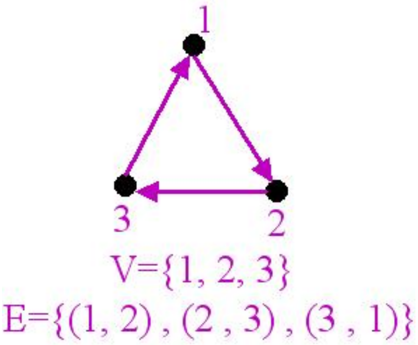
- For $k \geq 3$, use the preceding results with $n = 2^{\left\lfloor \frac{k}{2} \right\rfloor}$.
-

$$\binom{n}{k} 2^{1-\binom{k}{2}} < \frac{n^k}{k!} \cdot \frac{2^{1+k/2}}{2^{k^2/2}} = \left(\frac{2^{\left\lfloor \frac{k}{2} \right\rfloor}}{2^{k/2}} \right)^k \cdot \frac{2^{1+k/2}}{k!} \leq \frac{2^{1+k/2}}{k!} < 1$$

- For large values of k, $\dfrac{2^{1+k/2}}{k!} << 1$. Thus, we can present a coloring without a K_k by randomly coloring the edges.
  - If we wanted to present a coloring for K_{1024} without a K_{20}, a random coloring would be false with probability less than $\dfrac{2^{11}}{20!}$.

Tournaments and the property S_k

- *Tournament* on a set V of n players is an orientation $T = (V, E)$ of the edges of K_n.
- One of (x, y), (y, x) is in E but not both.
- Interpret edge (x, y) as x wins over y.
- T has property S_k if for *every* set of k elements there is one who beats them all.



The above triange has property S_1 but not S_2.

Existence of Tournaments with property S_k

- Erdős (1963) proved the existence of tournaments with property S_k.
- Consider a random tournament on n vertices.
  - Choose either (i, j) or (i, j) to belong in E.
- Claim: *If* $\binom{n}{k}(1 - 2^{-k})^{n-k} < 1$ *then there is a tournament on* n *vertices that has property* S_k.

Proof

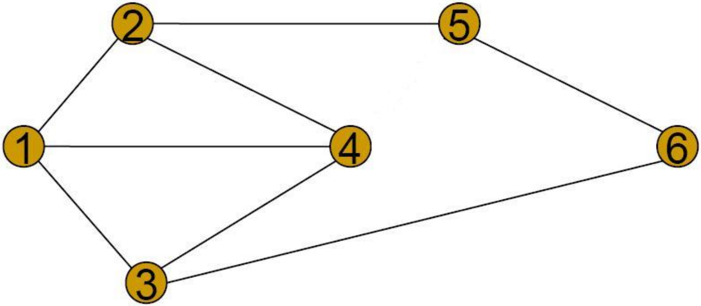
- Fix a set of k vertices K.
- Let A_K denote the event that there is no vertex of $V - K$ which beats all members of K.
- $\Pr[A_K] = (1 - 2^{-k})^{n-k}$.
- $$\Pr\left[\bigcup_K A_K\right] \leq \sum_K \Pr[A_K] = \binom{n}{k}(1 - 2^{-k})^{n-k} < 1$$
- The probability that no A_R occurs is positive, i.e. there is a tournament on n vertices that has property S_k.
- Using $\binom{n}{k} < \left(\frac{en}{k}\right)^k$ and $(1 - 2^{-k})^{n-k} < e^{-\frac{n-k}{k}}$ we have that

$$n_{\min} \leq k^2 \cdot 2^k \cdot \ln 2 \cdot (1 + o(1))$$

Much to know with so little effort!

Dominating sets

- A *dominating set* on a graph $G = (V, E)$ is a subset $U \subset V$ such that every vertex in $U - V$ has a neighbour in U.



In the above graph both $\{1, 6\}$ and $\{2, 3\}$ are dominating sets of size 2.

Dominating sets

- *Let* $G = (V, E)$ *a graph on* n *vertices, with minimum degree* $\delta > 1$. *Then* G *has a dominating set of at most* $n[1 + \ln(\delta + 1)]/(\delta + 1)$ *vertices.*

Proof:

- Let $p \in [0, 1]$ (to be determined later).
- Choose each vertex V into a set X independently with probability p.

Dominating sets

- Denote $Y = Y_X$ the set of vertices of $V - X$ that do not have a neighbour in X.
- $\Pr[v \in Y] \leq (1-p)^{\delta+1}$.
- If Y_i denotes the event that i belongsin Y, then $Y = \sum_1^n Y_i$ is the total number of vertices in Y.
- $E[|Y|] \leq n(1-p)^{\delta+1}$, $E[|X|] \leq np$.
- $E[|X| + |Y|] \leq np + n(1-p)^{\delta+1}$. Clearly $X \cup Y$ is a dominating set.
- There exists a dominating set of size at most $np + n(1-p)^{\delta+1}$!

Optimizing p

- Using $1 - p \leq e^{-p}$ we want to minimize the quantity

$$np + ne^{-p(d+1)}$$

- $p^* = \dfrac{\ln(\delta+1)}{\delta+1}$.
- Substituting back

$$|U| \leq np^* + ne^{-p^*(d+1)} = n \cdot \frac{\ln(\delta+1)}{\delta+1} + n \cdot \frac{1}{\delta+1}$$

$$|U| \leq n[1 + \ln(\delta+1)]/(\delta+1)$$

New techniques adapted

- Linearity of expectation. No need to worry about dependencies!
- If a variable has mean value m, then it *must* take at least one value $\leq m$ and at least one value $\geq m$.
- The optimization of the parameter p.

Finding the dominating set

We now present a greedy algorithm to obtain a dominating set of size at most $n[1+\ln(\delta+1)]/(\delta+1)$.

- For each vertex v denote by $C(v)$ the set consisting of v together with all its neighbors.
- Suppose that the number of vertices u that do not lie in the union of the sets $C(v)$ of the vertices chosen so far is r.
- The sum of the cardinalities of the sets $C(u)$ over all such uncovered vertices u is at least $r(\delta+1)$.
- Pick a vertex v that belongs to at least $r(\delta+1)/n$ such sets $C(u)$.
- The number of uncovered vertices is now at most $r(1-(\delta+1)/n)$.

Finding the dominating set

- The number of uncovered vertices decreases at each step by a factor of $1(\delta+1)/n$.
- After $n\ln(\delta+1)/(\delta+1)$ steps there will be at most

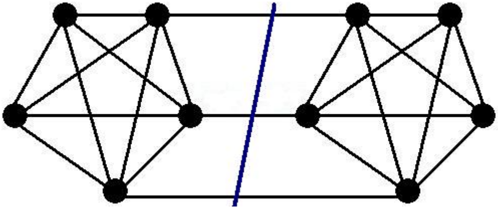
$$n\left(1-\frac{\delta+1}{n}\right)^{n\frac{\ln(\delta+1)}{\delta+1}} \leq ne^{-(\delta+1)\cdot\frac{\ln(\delta+1)}{\delta+1}} = \frac{n}{\delta+1}$$

  yet uncovered vertices
- Add these $n/(\delta+1)$ vertices to the dominating set to obtain the desired total

$$n\frac{\ln(\delta+1)}{\delta+1} + \frac{n}{\delta+1}$$

Determining $n/2$-edge connectivity

The *edge connectivity* of a graph G is the minimum size of a cut of G.

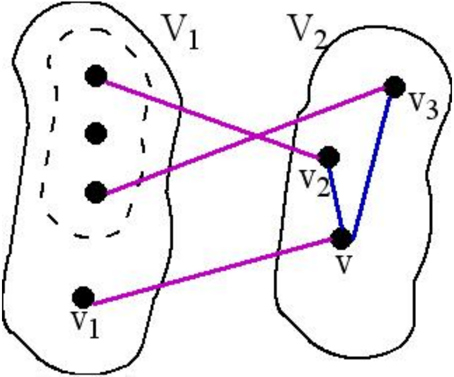


We will use the above ideas to determine if a graph is $n/2$-edge connected.

Lemma (Podderyugin and Matula)

Lemma. Let $G = (V, E)$ be a graph with minimum degree δ and let $V = V_1 \cup V_2$ be a cut of size smaller than δ in G. Then every dominating set U of G has vertices in V_1 and in V_2.

- Suppose $U \subset V_1$. Choose a vertex $v \in V_2$ and let $v_1, \ldots, v_\delta$ its neighbours.
- If $v_i \in V_1$, $e_i = \{v, v_i\}$.
- Else, there is a $u \in U$ such that $\{u, v_i\} \in E$. Choose $e_i = \{u, v_i\}$.
- e_i, $1 \leq i \leq \delta$ form a cut of size δ. Contradiction.



Algorithm to determine $n/2$-edge connectivity

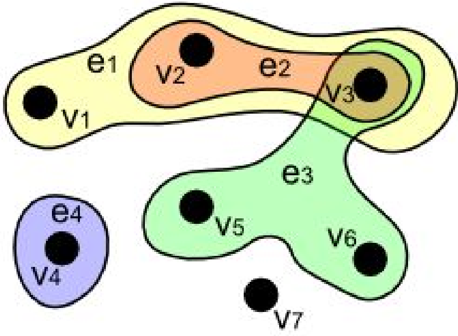
1. Check if the minimum degree δ of G is at least $n/2$.
   - If not, G is not $n/2$ edge-connected, and the algorithm ends.
2. There is a dominating set $U = \{u_1 \ldots, u_k\}$ of size at most

$$k \leq n\frac{\ln(\delta+1)+1}{\delta+1} \leq n\frac{\ln(n/2+1)+1}{n/2+1} = O(\log n)$$

   which can be found in $O(n^2)$.
3. For each i, $2 \leq i \leq k$, find the minimum size s_i of a cut that separates u_1 from u_i.
   - Solve a network flow problem in $O(n^{8/3})$.
4. By the previous lemma, the edge connectivity of G is the minimum between δ and $\min_{2\leq i\leq k} s_i$.
5. Total complexity $O(n^{8/3} \log n)$.

Hypergraphs

- A *hypergraph* is a pair $H = (V, E)$, where V is a finite set whose elements are called vertices and E is a family of subsets of V, called *edges*.
- It is *n-uniform* if each of its edges contains precisely n vertices.
- It is two-*colorable* if there is a two-coloring of V such that no edge is monochromatic.



Proving a lower bound

Let $m(n)$ denote the minimum possible number of edges of an n-uniform hypergraph that is not two-colourable.

- *Every n-uniform hypergraph with less than 2^{n-1} edges is two-colourable. Therefore $m(n) \geq 2^{n-1}$.*

Proof:

1. Consider such a graph and a random coloring of V with two colours.
2. For each edge $e \in E$, let A_e be the event that e is monochromatic. $\Pr[A_e] = 2^{1-n}$.
3. $$\Pr\left[\bigcup_e A_e\right] \leq \sum_e \Pr[A_e] < 2^{n-1} \cdot 2^{1-n} = 1$$
4. The result follows.

Proving an upper bound for $m(n)$

- Fix V with v points (to be determined later).
- χ a coloring of V with a vertices in one colour and $b = v - a$ vertices in the other.
- Choose randomly a n-subset S of V.
- $$\Pr[S \text{ is monochromatic under } \chi] = \frac{\binom{a}{n} + \binom{b}{n}}{\binom{v}{n}}$$
- Due to convexity of $\binom{y}{n}$

$$\Pr[S \text{ is monochromatic under } \chi] = \frac{\binom{a}{n} + \binom{b}{n}}{\binom{v}{n}} \geq p$$

where $p = 2\frac{\binom{v/2}{n}}{\binom{v}{n}}$.

Proving an upper bound for $m(n)$

- Choose m n-subsets $S_1, \ldots, S_m$ independently (m to be determined later).
- Let A_χ be the event that none of the S_i are monochromatic.

$$\Pr[A_\chi] \leq (1-p)^m$$

- Summing over all possible 2^v colorings

$$\Pr\left[\bigcup_\chi A_\chi\right] \leq 2^v(1-p)^m$$

- Choosing $m = \left\lceil \frac{v \ln 2}{p} \right\rceil$

$$2^v(1-p)^m \leq 2^v e^{-pm} < 2^v 2^{-v} = 1$$

and so $m(n) \leq m$

Proving an upper bound for $m(n)$

- It remains to optimize the ratio v/p.

$$p = 2\frac{\binom{v/2}{n}}{\binom{v}{n}} = 2^{1-n}\prod_{i=0}^{n-1}\frac{v-2i}{v-i} \sim 2^{1-n}e^{-n^2/2v}$$

- The optimal upper bound is

$$m(n) < (1+o(1))\frac{e\ln 2}{4}n^2 2^n$$

Sum-free sets

A set A is called *sum-free* if there are no $\alpha_1, \alpha_2, \alpha_3 \in A$ such that $a_1 + a_2 = a_3$.

Proposition [Erdős 1965]. *Every set $B = \{b_1, \ldots, b_n\}$ of n nonzero integers contains a sum-free subset A of size $|A| > n/3$.*

- If $p = 3k + 2$ is a prime
$$C = \{k + 1, \ldots, 2k + 1\}$$
is sum-free in $\mathbb{Z}_p$. Note that $\dfrac{|C|}{p-1} = \dfrac{k+1}{3k+1} > \dfrac{1}{3}$.
- Find such a p large enough and choose randomly $1 \leq x < p$.
- Define $d_i \equiv x b_i \pmod{p}$ for x chosen randomly in $[1, p]$.
- $\Pr[d_i \in C] = \dfrac{|C|}{p-1} > \dfrac{1}{3}$.
- Expected number of elements b_i such that $d_i \in C$ is more than $n/3$.

Sum-free sets

- Consequently, there is a x and a subset A of B with $|A| > n/3$, such that $xa (\mathrm{mod}\ p) \in C$ for all $a \in A$.
- A is sum-free.

Erdős-Ko-Rado Theorem

Definition. *A family $\mathcal{F}$ of sets is called intersecting if $A, B \in \mathcal{F}$ implies $A \cap B \neq \emptyset$, i.e. A, B share a common element.*

Suppose $n \geq 2k$ and let $\mathcal{F}$ be an intersecting family of k-element subsets of an n-set, for definiteness $\{0, \dots, n-1\}$.

Erdős-Ko-Rado Theorem. $|\mathcal{F}| \leq \binom{n-1}{k-1}$.

Lemma

Lemma. For $0 \leq s \leq n1$ set $A_s = \{s, s+1, \ldots, s+k-1\}$ where addition is modulo n. Then $\mathcal{F}$ can contain at most k of the sets A_s.

- Fix $A_s \in \mathcal{F}$.
- Consider the $2(k-2)$ sets A_t that intersect A_s.
- Pair them in $k-1$ pairs such that 2 sets in the same pair are disjoint.
- $\mathcal{F}$ may contain at most 1 set from each pair, so as to be intersecting.
- The lemma follows.

Proof (by Katona, 1972)

1. Choose randomly a permutation σ of $\{0, \ldots, n-1\}$ and independently a random $0 \leq i \leq n-1$.
2. Set $A = \{\sigma(i), \ldots, \sigma(k+i-1)\}$.
3. Conditioning on any choice of σ, the lemma gives
$$\Pr[A \in \mathcal{F}] \leq k/n.$$
4. A is in fact selected uniformly over all k-element sets, thus
$$\Pr[A \in \mathcal{F}] = \frac{|\mathcal{F}|}{\binom{n}{k}}$$
5. Combining,
$$|\mathcal{F}| \leq \frac{k}{n}\binom{n}{k} = \binom{n-1}{k-1}$$

Q.E.D