

Lecture 12: The Probabilistic Method

The probabilistic method is an elegant and powerful technique for analysis of algorithms and combinatorial structures. It is based on a simple premise: in order to prove the existence of a structure, we merely need to show that there is a positive probability that the structure exists.

This method was first used by Szele in 1943, but Erdős was the first to fully realize the power of this method and apply it to a wide variety of problems.

Chromatic Number of a Graph

A fundamental question in graph theory is the relationship between the subgraphs of a graph and its chromatic number. For example, can we show that triangle free graphs must have a chromatic number bounded by some constant k ?

Recall that a *proper vertex k -coloring* of a graph $G(V, E)$ is an assignment of k colors to vertices such that no two vertices of the same color share an edge. The *chromatic number* $\chi(G)$ is equal to the smallest number of colors needed to have a proper vertex coloring.

Definition: The *girth* of a graph G , $g(G)$, is the length of the smallest cycle in G .

In 1954, B. Descartes was the first to show that triangle-free graphs can have arbitrarily high chromatic numbers. But this construction contained many short cycles. Is there a dependence on short cycles and an arbitrarily high chromatic number?

In 1959, Paul Erdős used the probabilistic method to prove the existence of graphs with arbitrarily high girth and chromatic number.

Theorem 1 (Erdős, 1959) *For every $g, k > 0$, there exists a graph G with $\chi(G) \geq k$ and $g(G) \geq g$.*

Proof: Consider a random graph on n vertices $G(n, p)$ such that there is an edge between every pair of vertices with probability p independent of every other pair. The idea is that if a graph has chromatic number k , then there must exist at least one independent set of size at least $\frac{n}{k}$, since a coloring corresponds to an independent set.

Therefore, in order to show that $\chi(G) \geq k$, it suffices to prove that with high probability the size of any independent set in G is at most $\frac{n}{k}$. We will prove that with high probability, the graph doesn't have any independent set of size $\frac{n}{2k}$.

Recall from basic probability the “union bound” or “subadditivity property” of probabilities. That is, for any (arbitrary, not necessarily disjoint) events E_1, E_2, \dots, E_j ,

$$\Pr(\cup_{i=1}^j E_i) \leq \sum_{i=1}^j \Pr(E_i)$$

The probability that any set of $n/2k$ vertices is an independent set is $(1-p)^{\binom{n/2k}{2}}$. There are $\binom{n}{n/2k}$ possibilities for vertex sets of size $n/2k$. By the union bound, the probability of $G_{n,p}$ having such an independent set is therefore at most

$$\begin{aligned} \Pr \left[G_{n,p} \text{ has an independent set of size } \frac{n}{2k} \right] &\leq \binom{n}{n/2k} (1-p)^{\binom{n/2k}{2}} \\ &\leq 2^n e^{-pn^2/8k} \\ &\leq e^{\log 2n - pn^2/8k} \\ &\leq e^{n \log 2 - n^{\epsilon+1}/8k} \end{aligned}$$

where we set $p = n^{\epsilon-1}$ for some $\epsilon < 1$. The above expression tends to zero as $n \rightarrow \infty$, so is therefore smaller than $1/4$ for n large enough.

Let X be the random variable counting the number of cycles of length g and smaller. By linearity of expectation,

$$\begin{aligned} \mathbb{E}[X] &= \sum_{i=1}^g \binom{n}{i} \frac{(i-1)!}{2} p^i \\ &\leq g(np)^g = gn^{\epsilon g} \end{aligned}$$

if $0 < \epsilon < 1/g$, the above expression is $o(n)$. Thus, for sufficiently large n , $\mathbb{E}(X) < 1/4$. By Markov's inequality,

$$\Pr(X > n/2) \leq \Pr(X > 2\mathbb{E}(X)) < 1/2$$

Therefore, if we choose n sufficiently large, and $p = n^{\epsilon-1}$ for $0 < \epsilon < 1/g$, the probability that $G_{n,p}$ has a independent set of size $\frac{n}{2k}$ or that the number of cycles of length at most g is $\frac{n}{2}$ is less than 1 by the union bound.

Now, we can construct a graph G' by removing a vertex from each short cycle. The number of vertices in G' is at least $\frac{n}{2}$, the size of the maximum independent set in G' is no more than $\frac{n}{2k}$, and there are no cycles of length less than g . This implies that $\chi(G') > k$ and $g(G') > g$.

■