I would like to demonstrate one proof of a result of Sauer, Perles and Shelah, Vapnik and Chervonenkis from 1971,1972. The proof is due to Smolensky and is from 1997. There are a variety of proofs, basic induction works fine.

The subject of Extremal Combinatorics considers the maximum number of objects you can find subect to some conditions (and if that is possible then considers structures you would encounter). If we consider subsets of $\{1, 2, \ldots, m\}$, then there are at most 2^m subsets. We will encode a family of sn ubsets of $\{1, 2, \ldots, m\}$ as an $m \times n$ matrix $A = (a_{ij})$ with entries 0,1 and for which the jth column of A, which we can denote as A_j, corresponds to the jth subset A_j and we have $a_{ij} = \begin{cases} 1 & \text{if } i \in A_j \\ 0 & \text{if } i \notin A_j \end{cases}$ We will be asking how many columns n can such a matrix have subject to some restriction involving submatrices.

First consider the vector space of polynomials. We say a polynomial in variables $x_1, x_2, \ldots, x_m$ is *multilinear* of it has no expressions containing x_i^t for $t \geq 2$. Thus it is linear in each variable (when considering the other variables fixed). The degree of such a polynomial is given by the usual definition. Define

$$V = \{\text{multilinear polynomials of degree} \leq 2\}$$

Then $\dim(V) = \binom{m}{2} + \binom{m}{1} + \binom{m}{0}$ since we readily find a basis $\{1, x_1, x_2, \ldots, x_m, x_1x_2, x_1x_3, \ldots, x_{m-1}x_m\}$ of size $\binom{m}{2} + \binom{m}{1} + \binom{m}{0}$.

Our object of study are so called *simple* matrices whose entries are either 0 or 1 with the additional condition that no column is repeated. As we have seen above, these are like families of subsets of a given set. Thus if A is an $m \times n$ simple matrix then it is 'easy' to see that $n \leq 2^m$ because there are only 2^m possible columns of 0's and 1's. We wish to impose an additional property on A and obtain a good bound on n as a function of m. Let

$$K_3 = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

We want A not to contain K_3 as a *configuration*, namely there is no 3×8 submatrix of A which is a row and column permutation of K_3. Alternatively, if we form a 3×8 matrix by deleting 2 rows of A and all but 8, then the result should not consist of all possible (0,1)-columns on 3 rows.

The following 5×16 matrix has the desired property.

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

This follows because each triple of rows of the above matrix avoids

$$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

Since any row and column permutation of K_3 would contain this 3×1 column, we deduce that our 5×16 matrix avoids K_3. There are many ways to create a matrix avoiding K_3 and you should not expect them to be so symmetric. This was just a quick example.

The following theorem has a remarkable number of applications. Google VC-dimension and you'll find some references to Learning Theory and other topics. The following does generalize to K_k but for definiteness we only consider K_3.

Theorem 1 *(Vapnik and Chervonenkis 1971, Sauer 1972, Perles and Shelah 1972)*
Let A be an $m \times n$ simple matrix with no configuration K_3 (with no 3×8 submatrix which is a row and column permutation of K_3). Then

$$n \le \binom{m}{2} + \binom{m}{1} + \binom{m}{0}.$$

Proof: To prove this lovely result we need some polynomials, one per column, that are linearly independent and are multilinear and of degree at most 2. If we achieve this, then we will have proven the bound. We will *only* be evaluating these polynomials on the values of columns of A but that will suffice to have them be linearly independent.

For the ith column A_i of A we create a polynomial as follows. Let $A_i = (a_1, a_2, \ldots, a_m)^T$. We form a polynomial in $\mathbf{x} = (x_1, x_2, \ldots, x_m)^T$

$$f_i(\mathbf{x}) = \prod_{j=1}^{m} (1 - x_j - a_j)$$

Thus $f_i(A_i) \ne 0$ while $f_i(A_k) = 0$ for $k \ne i$. Certainly this means the polynomials are linearly independent. Also they are multilinear. But unfortunately they have degree m. So we have some work to do!

We will only be evaluating these polynomials on columns of A so that $x_i \in \{0, 1\}$. As well $a_i \in \{0, 1\}$. We note that $(1 - x_i - a_i) = 0$ for $x_i \ne a_i$, $(1 - x_i - a_i) = 1$ for $x_i = a_i = 0$ and $(1 - x_i - a_i) = -1$ for $x_i = a_i = 1$.

Now we use the fact that A has no configuration K_3. For each triple of rows i, j, k there must be some column of three elements missing say perhaps

$$\text{no} \quad \begin{array}{c} i \\ j \\ k \end{array} \begin{bmatrix} c \\ d \\ e \end{bmatrix}$$

We can form a polynomial

$$f_{ijk}(\mathbf{x}) = (1 - x_i - c)(1 - x_j - d)(1 - x_k - e)$$

which has the property that evaluated at any column $\mathbf{y}$ of A, $f_{ijk}(\mathbf{y}) = 0$. Now $f_{ijk} = -x_i x_j x_k + (1-e)(x_i x_j) + (1-d) x_i x_k + (1-c) x_j x_k - (1-c)(1-d) x_k - (1-c)(1-e) x_j - (1-d)(1-e) x_i + (1-c)(1-d)(1-e)$. Using $f_{ijk}(\mathbf{y}) = 0$ we can use the identity

$$\begin{aligned} x_i x_j x_k = {} & (1-e)(x_i x_j) + (1-d) x_i x_k + (1-c) x_j x_k - (1-c)(1-d) x_k \\ & -(1-c)(1-e) x_j - (1-d)(1-e) x_i + (1-c)(1-d)(1-e), \end{aligned}$$

at least when evaluated on the columns of A.

The right hand side has terms of degree at most 2. We use such an identity on our polynomials f_ℓ for each ℓ, taking any term that contains the product $x_i x_j x_k$ and replacing $x_i x_j x_k$ by $(1-e)(x_i x_j) + (1-d) x_i x_k + (1-c) x_j x_k - (1-c)(1-d) x_k - (1-c)(1-e) x_j - (1-d)(1-e) x_i + (1-c)(1-d)(1-e)$. Repeat over and over again to get a polynomial $f'_\ell(\mathbf{x})$ that agrees with $f_\ell(\mathbf{x})$ on columns of A

and has degree at most 2. The n polynomials $f'_\ell(\mathbf{x})$ are linearly independent and all in V (still multilinear). Now $\dim(V) = \binom{m}{2} + \binom{m}{1} + \binom{m}{0}$ and so we conclude

$$n \leq \binom{m}{2} + \binom{m}{1} + \binom{m}{0}. \qquad \blacksquare$$