组合数学第十五讲

授课时间: 2018年12月28日　授课教师: 孙晓明

1 不相同子集和问题的进一步讨论

定理 1. 从$\{1, 2, \ldots, 100\}$的任意9元子集中，必能找到两个元素和相同的不相交非空子集。

证明 (此证明来自张家琳老师). 考虑任意一个9元集$\{a_1, a_2, ..., a_9\} \subseteq [100]$，其中$a_1 < a_2 < ... < a_9$。这个9元集包含的所有4元，5元，6元子集共有$\binom{9}{4} + \binom{9}{5} + \binom{9}{6} = 336$个。

令$k = a_3 + a_4 + a_5$，$S = \{a_1, a_3, a_4, a_5\}$，$T = \{a_3, a_4, a_5, a_7, a_8, a_9\}$。$S$内所有元素的和至少为$k+1$，$T$内所有元素的和至多为$k + 297$。考虑所有内部元素和比$S$小的$4, 5, 6$元集，其必为4元集且一定包含$a_1, a_2$。这样的集合至多有$\binom{7}{2} = 21$个。考虑所有内部元素和比$T$大的$4, 5, 6$元集，其一定包含元素$a_6, a_7, a_8, a_8$。这样的集合至多有$\binom{5}{2} + \binom{5}{1} + \binom{5}{0} = 16$个。

因此，至少有$336 - 21 - 16 = 299$个$4, 5, 6$元集，其内部元素和在区间$[k + 1, k + 297]$内。注意到这个区间内一共有297个整数，由鸽笼原理，一定有两个子集，它们的内部元素和相同。这就完成了定理的证明。 □

2 专题：拉姆塞理论（Ramsey Theory）

先来考虑这样一个简单的例子：

例1 从6个人中，一定能找到这样的3个人：他们要么两两相互都认识，要么两两相互都不认识。

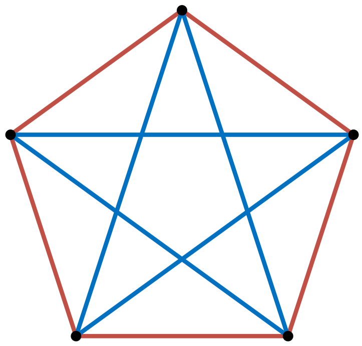
用图论的语言，上述结论可以被重述为：对6阶完全图K_6的所有边任意红蓝二染色，必存在同色三角形。其证明如下：

证明 随意选取一个顶点v_1，从这个点共引出了5条边。根据鸽巢原理，其中必存在三条边颜色相同，不妨设这三条边被染成红色，又记这三条边连到的顶点为v_2, v_3, v_4，如这三个顶点之间存在红色边，那么与这条红色边相连的两个顶点与v_1组成了一个红色三角形；如这三个顶点之间无红色边，则v_2, v_3, v_4是一个蓝色三角形，原命题得证。 □

我们可以把上述问题推广到一般情况，为了方便描述，做如下定义：

定义2 (拉姆塞数). 定义拉姆塞数$R(n, m)$为使得如下条件成立的最小整数l：对l阶完全图K_l的所有边做任意红蓝二染色，其中要么存在n个点，它们之间的边均为红色；要么存在m个点，它们之间的边均为蓝色。

例1说明了$R(3, 3) \leq 6$，而对于K_5，存在如下反例：



其中显然不存在蓝色或红色的K_3子图，从而有$R(3,3)>5$，进而得到$R(3,3)=6$。

本质上，拉姆塞理论是鸽巢原理的推广，即当某个系统足够大时，必然能从中找到某种具有特定性质的子结构。关于拉姆塞数，我们还知道如下平凡结论：

1) $R(2,m)=m$，其中所谓红色的K_2就是一条红边，而从不存在红边的K_m中，显然可以找到一个蓝色的K_m；

2) $R(n,m)=R(m,n)$，一方面，考虑完全图的某种染色，将其所有边的颜色改变，得到的染色称为这种染色的补染色，那么K_l的全体染色的补染色显然包含了所有染色的情况；另一方面如果在某种染色中可以找到红色的K_n或蓝色的K_m，那么在它的补染色中一定可找到红色的K_m或蓝色的K_n，综上，推知拉姆塞数是对称的。

$R(3,3)=6$说明在K_6中必存在纯色三角形，可以证明，任意K_6的二染色中至少存在两个同色三角形。

计算拉姆塞数是一类很难的问题，只有很少的拉姆塞数被确切的算出。因此，我们希望先分析拉姆塞数的增长趋势，以下定理给出了$R(n,m)$的一个上界。

例2 对任意的$n,m>1$，有$R(n,m)\le R(n-1,m)+R(n,m-1)$。

证明 考虑对$R(n,m-1)+R(n-1,m)$阶完全图的任意红蓝二染色，其中某个节点v引出的$R(n,m-1)+R(n-1,m)-1$条边被任意染色，这些边中要么有$R(n,m-1)$条蓝色边，要么有$R(n-1,m)$条红色边。不失一般性地，假设存在与v相连的$R(n,m-1)$条蓝色边，那么考虑这些边另一端的$R(n,m-1)$个顶点的导出子图，其中要么存在蓝色的K_{m-1}，与v组成蓝色边的m阶团，要么存在红色边的n阶团，即：

$$R(n,m)\le R(n-1,m)+R(n,m-1)$$

□

发现拉姆塞数上界的累加方式与组合数非常相似，且易知：

$$R(n,1)=R(1,n)\le\binom{n+1}{1}=\binom{n+1}{n} \tag{2}$$

归纳假设$n+m<t$时，不等式$R(n,m)\le\binom{n+m}{n}$成立。当$n+m=t$时，如$\min\{n,m\}=1$，由(2)式推知归纳成立，否则根据归纳假设，有：

$$\begin{aligned}R(n,m)&\le R(n-1,m)+R(n,m-1)\\&\le\binom{n+m-1}{n-1}+\binom{n+m-1}{n}\\&=\binom{n+m}{n}\end{aligned}$$

归纳成立，证得$R(n,m)\le\binom{n+m}{n}$。

经过更精细的分析，该上界可以略微改进到$R(n,m)\le\binom{n+m-2}{n-1}$，与已知的最好下界距离很远，换言之目前我们甚至无法近似地描述拉姆塞数的增长趋势。

3 概率方法（Probabilistic Methods）与拉姆塞数下界

3.1 构造性的下界证明

回顾上节课，我们定义了红蓝二染色情形下的拉姆塞数$R(s,t)$，它表示满足下列条件的n的最小值：*将完全图K_n的每一条边任意涂成红色或蓝色，则在K_n中一定存在红色的K_s或蓝色的K_t。* 对于拉姆塞数表，我们更关心对角线上的值，即$s=t=n$时的情形。实际上，已知的拉姆塞数非常少，对角线上的拉姆塞数目前也只能求解到$R(4,4)$，下面我们将对拉姆塞数$R(n,n)$的阶进行估计。上节课我们已经得到了$R(n,n)$的一个上界：

$$R(n,n)\le\binom{2n}{n}\sim\frac{2^{2n}}{\sqrt{n}}$$

下面我们来求解$R(n,n)$的下界，即找到一个尽可能大的N，使得K_N的某种二染色中既没有红色K_n也没有蓝色K_n。先来看$n=6$的情况：

定理 3. $R(6,6)>25$。

证明 我们将给出25阶完全图的一个染色方案。首先，将25个点每组5个分为5组，将同组中的点两两用蓝边相连，不同组之间的点两两用红边相连。根据鸽巢原理，任意取6个点必存在两个点落在同一组中，故任意6个点的导出子图必包含蓝边，即该染色方案下，没有红色的K_6；同时，必有两点在不同组中，故导出子图中必有红边，推知该染色方案下没有蓝色的K_6。 □

同理可证$R(n,n)\ge(n-1)^2+1=\Omega(n^2)$，而$R(n,n)$的上界却是指数量级；目前最好的构造性证明得出的下界为$R(n,n)=\Omega(2^{C\log^2 n})$，其中$C$为某个常数，仍然和上界相差很大。下面我们将通过概率方法（Probabilistic Methods），存在性地证明一个远好于此的下界。

3.2 概率方法与存在性的下界证明

概率方法是由保罗·埃尔德什（Paul Erdös）发展出的一套证明技术，他率先用这种方法将$R(n,n)$的下界提高到了指数量级，现在我们用概率方法来证明下面的定理。

定理 4. *对于任意正整数n，其拉姆塞数$R(n,n)$满足：*

$$R(n,n)=\Omega(2^{n/2})$$

证明上述定理之前，我们需要先证明如下引理：

引理5 (Union bound). 对于任意的n个随机事件$A_1, A_2, \ldots, A_n$，如下不等式成立：

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} \Pr(A_i)$$

证明　当n为1时引理显然成立，归纳假设当$n < t$时引理成立，当$n = t$时，有：

$$\begin{aligned}
\Pr\left(\bigcup_{i=1}^{n} A_i\right) &= \Pr\left(A_1 \cup \bigcup_{i=2}^{n} A_i\right) \\
&= \Pr(A_1) + \Pr\left(\bigcup_{i=2}^{n} A_i\right) - \Pr\left(A_1 \cap \bigcup_{i=2}^{n} A_i\right) \\
&\le \Pr(A_1) + \Pr\left(\bigcup_{i=2}^{n} A_i\right) \\
&\le \sum_{i=1}^{n} \Pr(A_i)
\end{aligned}$$

归纳成立，原引理得证。 □

Union bound是概率方法中用到的主要工具之一，借助Union bound我们可以证明定理2：

定理2的证明　下面证明$R(n,n)$的下界：

$$R(n,n) > \frac{2^{n/2} \cdot n}{100}$$

记$N := \frac{2^{n/2}\cdot n}{100}$，证明该下界也就是要证明存在$K_N$的某种红蓝二染色方案，其中既没有红色$K_n$，也没有蓝色$K_n$，我们将符合上述条件的染色称为“好的染色方案”；相反地，将那些存在红色K_n 或者蓝色K_n的染色方案称为“坏的染色方案”。现对K_N中所有的$\binom{N}{2}$条边独立随机地做二染色，形成随机染色图G。利用古典概型的性质，我们有：

$$\Pr(G\text{是坏的}) = 1 - \Pr(G\text{是好的})$$

由Union bound得到：

$$\begin{aligned}
\Pr(G\text{是坏的}) &\le \Pr(\text{存在红色}K_n) + \Pr(\text{存在蓝色}K_n) \\
&= 2 \cdot \Pr(\text{存在红色}K_n)
\end{aligned}$$

进一步，根据Union bound有如下推导：

$$\begin{aligned}
\Pr(G\text{是坏的}) &\le 2 \cdot \Pr\left(\bigcup_{1 \le i_1 < i_2 < \cdots < i_n \le N} (v_{i_1}, v_{i_2}, \ldots, v_{i_n}\text{构成红色}K_n)\right) \\
&\le 2 \cdot \sum_{1 \le i_1 < i_2 < \cdots < i_n \le N} \Pr\left(v_{i_1}, v_{i_2}, \ldots, v_{i_n}\text{构成红色}K_n\right) \\
&= 2 \cdot \binom{N}{n} \cdot \left(\frac{1}{2}\right)^{\binom{n}{2}}
\end{aligned}$$

已知$\binom{n}{m} \le (\frac{ne}{m})^m$，据此对上式进行放缩：

$$\begin{aligned}\Pr(G\text{是坏的}) &\le 2\cdot\binom{N}{n}\cdot\left(\frac{1}{2}\right)^{\binom{n}{2}}\\ &\le 2\cdot\left(\frac{2^{n/2}\cdot n\cdot e}{100n}\right)^n\cdot\left(\frac{1}{2}\right)^{\frac{n(n-1)}{2}}\\ &\le 2\cdot\left(\frac{\sqrt{2}e}{100}\right)^n < 1\end{aligned}$$

推知出现好的染色方案的概率不为零，即在所有的染色中均存在好的染色方案。 □

概率方法的应用非常广泛，我们继续用它证明如下定理：

定理 6. 对于一个图$G=(V,E)$，它的一个点集划分是满足$A\cup B=V, A\cap B=\phi$的一个顶点子集对(A,B)，由点集划分(A,B)导出的一个边集$[A,B]=\{e=(a,b)\in E|a\in A, b\in B\}$称为$G$的一个割。求证：对任意图$G=(V,E)$，存在一个割包含至少$\frac{|E|}{2}$条边。

证明 考虑如下算法：初始状态两个集合$A=B=\phi$。对于图G中的每个顶点v，抛一枚正反面概率均为$\frac{1}{2}$的硬币。若硬币正面则将v加入A，否则将v加入B。当所有顶点都有归属时算法终止，输出$[A,B]$。

现在对这个算法效率进行分析。容易验证对任意$e\in E$，它的两个顶点分属A和B的概率为$\frac{1}{2}$。对每条边$e\in E$，设随机变量X_e满足若$e\in[A,B]$则$X_e=1$，否则$X_e=0$。则有$\mathrm{E}[X_e]=\frac{1}{2}$对任意边$e$成立。设随机变量$X$表示之前算法得到的割中的边数，则$\mathrm{E}[X]=\sum_{e\in E}\mathrm{E}[X_e]=\frac{|E|}{2}$。由此可以得到$\Pr(X\ge\frac{|E|}{2})>0$，故图$G$存在一个割包含至少$\frac{|E|}{2}$条边。 □