四色问题——平凡问题中的复杂

不久前上映的《嫌疑人 X 的献身》里出现的四色问题一定给大家留下了很深刻的印象。那，我们一起来记录细节吧

定理概要

提出者

Francis Guthrie

提出时间

1852 年

应用学科

拓扑学、图论

适用领域范围

排程和分配问题，~~地图编辑~~

地图绘制并不需要四色定理：他只要着色，不需要用最少的颜色。实际画地图时一般不用四种颜色

四色问题又称四色猜想、四色定理，是世界近代三大数学猜想之一（世界三大数学猜想即费马猜想、四色猜想和哥德巴赫猜想）。地图四色定理（Four color theorem）最先是由一位叫古德里（Francis Guthrie）的英国大学生提出来的。

四色问题的内容是“任何一张地图只用四种颜色就能使具有共同边界的国家着上不同的颜色。”也就是说在不引起混淆的情况下一张地图只需四种颜色来标记就行。

用数学语言表示即“将平面任意地细分为不相重叠的区域，每一个区域总可以用 1234 这四个数字之一来标记而不会使相邻的两个区域得到相同的数字。”这里所指的相邻区域是指有一整段边界是公共的。如果两个区域只相遇于一点或有限多点就不叫相邻的。因为用相同的颜色给它们着色不会引起混淆。

发展简史

•问题的提出

1852 年，毕业于伦敦大学的格斯里（FrancisGuthrie）来到一家科研单位搞地图着色工作时，发现每幅地图都可以只用四种颜色着色。这个现象能不能从数学上加以严格证明呢？他和他正在读大学的弟弟决心试一试，但是稿纸已经堆了一大叠，研究工作却是没有任何进展。

1852 年 10 月 23 日，他的弟弟就这个问题的证明请教了他的老师、著名数学家德·摩尔根，摩尔根也没有能找到解决这个问题的途径，于是写信向自己的好友、著名数学家哈密顿爵士请教，但直到 1865 年哈密顿逝世为止，问题也没有能够解决。

1872 年，英国当时最著名的数学家凯利正式向伦敦数学学会提出了这个问题，于是四色猜想成了世界数学界关注的问题，世界上许多一流的数学家都纷纷参加了四色猜想的大会战。

从此，这个问题在一些人中间传来传去，当时，三等分角和化圆为方问题已在社会上“臭名昭著”，而“四色瘟疫”又悄悄地传播开来了。

•肯普的研究

1878～1880 年两年间，著名的律师兼数学家肯普(Alfred Kempe)和泰勒(Peter Guthrie Tait)两人分别提交了证明四色猜想的论文，宣布证明了四色定理。

大家都认为四色猜想从此也就解决了，但其实肯普并没 v 有证明四色问题。11 年后，即 1890 年，在牛津大学就读的年仅 29 岁的赫伍德以自己的精确计算指出了肯普在证明上的漏洞。他指出肯普说没有极小五色地图能有一国具有五个邻国的理由有破绽。不久泰勒的证明也

被人们否定了。人们发现他们实际上证明了一个较弱的命题——五色定理。就是说对地图着色，用五种颜色就够了。

不过，让数学家感到欣慰的是，郝伍德没有彻底否定肯普论文的价值，运用肯普发明的方法，郝伍德证明了较弱的五色定理。这等于打了肯普一记闷棍，又将其表扬一番，总的来说是贬大于褒。真不知可怜的肯普律师是什么心情。 追根究底是数学家的本性。一方面，五种颜色已足够，另一方面，确实有例子表明三种颜色不够。那么四种颜色到底够不够呢？这就像一个淘金者，明明知道某处有许多金矿，结果却只挖出一块银子，你说他愿意就这样回去吗？

·肯普的错误中的贡献

肯普是用归谬法来证明的，大意是如果有一张正规的五色地图，就会存在一张国数最少的“极小正规五色地图”，如果极小正规五色地图中有一个国家的邻国数少于六个，就会存在一张国数较少的正规地图仍为五色的，这样一来就不会有极小五色地图的国数，也就不存在正规五色地图了。这样肯普就认为他已经证明了“四色问题”，但是后来人们发现他错了。

四色猜想在短短的两年时间里被一个并非“专业”数学家的“外行人”解决，让很多当初认为这个问题是难题的数学家觉得，这个问题也许并没有涉及到数学中深层的本质难点。对四色问题的研究逐渐减少，数学家们已经将其视为事实。刘易斯·卡罗尔将四色问题化为游戏：一方设计地图，另一方来为其着色。1886 年，英国男校克里夫顿学院（德语：Clifton College）校长将四色问题作为给全校学生挑战的难题，要求答案长度“不得超过一页纸的文字，30 行算式以及一页纸的图”。

德国数学家菲利克斯·克莱因甚至将这个问题和 1840 年莫比乌斯提出并解决的另一个问题相混淆起来，认为四色问题不过是后者的直接推论。这个误解被几何学家理查德·巴尔策（德语：Heinrich Richard Baltzer）在 1885 年重复，导致直到 21 世纪仍有类似的传言。而实际上莫比乌斯解决的是完全图K_5 ，不是平面图的问题，与四色问题没有直接联系。

然而，在肯普的证明发表的 11 年之后，珀西·约翰·希伍德（英语：Percy John Heawood）发表一篇文章，指出肯普的证明中包含一个错误。希伍德在文章中遗憾地指出，他无法修正这个错误，以得到一个四色问题的正确证明，因此他的文章更多是摧毁而非建设（rather destructive than constructive）。不过，尽管无法得到四色定理，希伍德仍然在肯普的思路上前进，得到一个较弱的定理：五色定理。

根据希伍德的说明，肯普的错误在于证明 5 邻国是可约构形时，构造两条肯普链以换色，然而第二次换色时，肯普的方法并不总是成功的。希伍德提供一个包含 25 个国家的地图作为反例。

希伍德的报告是由肯普自己提交给伦敦皇家数学学会的。肯普承认自己的证明中存在缺陷，并且他未能去除这个缺陷。然而希伍德的工作并没有受到应有的重视。数学界普遍认为这只是无关紧要的错误，很快就能得到纠正。1894 年创刊的《L'intermédiaire des mathématiciens》杂志以四色问题作为头一个征解问题，结果很快就收到解答，称其已被解决，并引用了肯普、泰特等人的论文。E.吕卡的《娱乐数学》（Récréations mathématiques）第四卷提到肯普的证明，但丝毫没提到希伍德已经指出肯普证明的错处。

直到世纪之交时，数学家们仍旧认为，四色问题所需要的只是某个灵光一现的妙想。一个广为流传的故事是闵可夫斯基在教拓扑学课时提到四色问题，说：“这个问题一直没有解决，只是因为试图解决它的都是三流的数学家”。他声称要在课上证明之，但直到下课仍然无法成功，在耗费若干堂课的时间后，只能承认失败

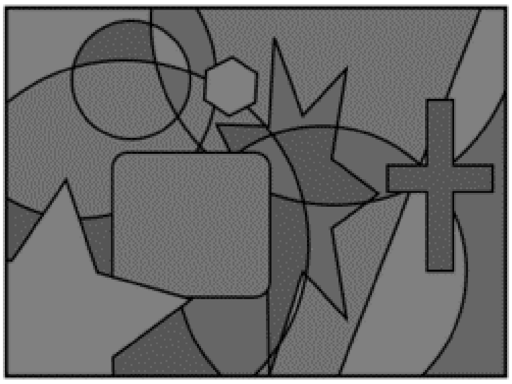
不过肯普的证明阐明了两个重要的概念，对以后问题的解决提供了途径。**第一个概念是“构形”**。他证明了在每一张正规地图中至少有一国具有两个、三个、四个或五个邻国，不存在每个国家都有六个或更多个邻国的正规地图，也就是说，由两个邻国，三个邻国、四个或五个邻国组成的一组“构形”是不可避免的，每张地图至少含有这四种构形中的一个。

肯普提出的另一个概念是“可约”性。“可约”这个词的使用是来自肯普的论证。他证明了只要五色地图中有一国具有四个邻国，就会有国数减少的五色地图。自从引入“构形”，“可约”概念后，逐步发展了检查构形以决定是否可约的一些标准方法，能够寻求可约构形的不可避免组，是证明“四色问题”的重要依据。但要证明大的构形可约，需要检查大量的细节，这是相当复杂的。

·缓慢的进展

人们发现四色问题出人意料地异常困难，曾经有许多人发表四色问题的证明或反例，但都被证实是错误的。后来，越来越多的数学家虽然对此绞尽脑汁，但一无所获。于是，人们开始认识到，这个貌似容易的题目，其实是一个可与费马猜想相媲美的难题。　进入 20 世纪以来，科学家们对四色猜想的证明基本上是按照肯普的想法在进行。

四色定理的本质就是在平面或者球面无法构造有五个或者五个以上的两两相连的区域，如果有五个以上两两相连区域，第五个区域至少与一个区域同一种颜色。这个理论在其他构造中是显然的，例如在环面上（亏格为 1），需要 7 色，就是因为环面不能构造 8 个两两相连区域。在亏格为 2 的双环面上，需要 8 色，就是不能构造 9 个区域两两相连。



20 世纪起，欧洲数学界对四色定理的研究出现停滞。相反地，这个问题在美国得到更多的关注。不少杰出的数学家研究了这个问题，并作出很大贡献。一部分的努力是修正肯普的证明；另一方面的努力是延续泰特的思路，将四色问题进行转化，以使用更多有力的数学工具。

对四色问题的转化在泰特之后并未停止过。从拓扑学的版本转化至图染色的版本后，希伍德又在 1898 年提出新的变形。肯普和泰特已经注意到，证明四色问题只需要考虑三个国家有共同“交点”的情况，更多国家有共同交点的情形可以转化为前者。因此这样对应的染色图中，每个顶点恰会连出三条边。这样的图被称为“三度图”（trivalent map）。希伍德观察到，如果三度图中任意由边围成的区域，边的个数都是 3 的倍数，那么图可以被 4-染色。他进一步发现，只要存在一种给图的顶点赋值+1 或-1 的方法，使得每个区域的顶点数字之和都被 3 整除，那么图可以被 4-染色。可以证明，4-染色和存在赋值方法是等价的。

在美国，数学家对四色定理的研究从未停止过。除了约翰·霍普金斯大学的皮尔斯以及斯多利等人外，另一个研究者是保罗·温尼克（英语：Paul Wernicke）。从当时的学术圣地哥廷根大学毕业的温尼克来到美国后在肯塔基大学任教。他 1904 年发表的论文中已经出现了可约性的雏形。然而美国数学界在四色问题上首次实质性的进展出现在 1912 年后。普林斯顿大学的奥斯瓦尔德·维布伦（经济学家托尔斯坦·范伯伦的侄子）是这波浪潮的先锋。他的工作重心是拓扑学，1905 年证明了若尔当曲线定理。对庞加莱发展出的新代数工具有深入了解的他，很自然地开始对四色定理的研究。他使用有限几何学的观念和有限域上的关联矩阵（英语：incidence matrix）作为工具，将四色问题转化成有限域系数空间上的方程问题。这个方向被后来的密码学家、数学家威廉·托马斯·塔特（英语：William Thomas Tutte）称为“量化方法”（the quantitative method）。同年，他的普林斯顿同僚乔治·戴维·伯克霍夫也开始探索这个方向，但一年之后他开始转向肯普的方法，也即是塔特所称的“定性方法”（the qualitative method），并提出可约环（reducible ring）的概念。1913 年，伯克霍夫发表名为《地图的可约性》（The Reducibility of Maps）的论文，利用可约环证明了：由不超过 12 个国家构成的地图都能用四色染色。1922 年，伯克霍夫的学生菲利普·富兰克林（英语：Philip Franklin）运用同样的方法，将结论加强到：不超过 25 个国家构成的地图都能用四色染色。**由于别克霍夫首次证明四色定理对不超过 12 个国家的地图成立，历史上证明的可染色地图的国家数上限记录被称为别克霍夫数。**

伯克霍夫等人的证明是肯普的方法的延续和系统化，归纳为寻找一个不可避免的可约构形集（an unavoidable set of reducible configurations）。这个理念已经体现在肯普的证明中。他首先说明任一地图中必然存在以下四种构形：2 邻国国家、3 邻国国家、4 邻国国家和 5 邻国国家；然后证明每种构形都是可约构形。后来希尔将这种分类方式称为“不可避免集”。伯克霍夫的构想是使用反证法：反设存在至少需要五种颜色染色的地图，那么其中必然存在国家数最小的“极小五色地图”（five-chromatic map）。这个地图必然是“不可约的”（irreducible）。而只要找到一组构形，使极小五色地图中不可避免地会出现其中一种构形，并且每个构形都是可约的，那么就能够通过约化，将地图的国家数减少，从而导致矛盾

肯普找的不可避免集由四种构形组成，但他无法证明最后一种（5 邻国国家）的可约性，因此伯克霍夫开始寻找刻画不可避免集的新方法。他提出以相邻国家连成的环来将整个地图 M 分为三个部分：环内部分 A、环外部分 B 以及环本身 R。若环上的国家数为 n 就称其为 n-环。如果 R 的任意染色都不妨碍 A 进行染色，那么就可以“忽略”A 而将 M 的染色问题约化为 B+R 的染色问题。这时便称 A+R 是可约构形，R 称为可约环。伯克霍夫证明了：当 R 是 4-环，或者 R 是 5-环且 A 中国家不止一个，或者 A+R 是“伯克霍夫菱形”时，A+R 都是可约的构形。因此极小五色地图不可能包含这些构形。富兰克林进一步证明：极小五色地图中必定包含三个邻接的五边国（5 邻国的国家），或者邻接的两个五边国与一个六边国，或者邻接的一个五边国和两个六边国。他从而得出一系列的可约构形，形成了 25 国以下地图的不可避免的可约构形集。因此推出，极小五色地图必定至少包含 26 个国家

这种方法的终极目标是找到所有地图的不可避免的可约构形集。然而随着国家数增多，要找到不可避免集并证明其可约化性就越难。这主要是因为随着环的增大，染色的方法数目会迅速增大。6-环的 4-染色方法有 31 种，而 12-环则有 22144 种。因此对大环围成的构形验证可约

性是十分繁杂的工作。1926 年，C.N.Reynolds 将别克霍夫数从 25 提高到 27。1938 年，富兰克林将其推进到 31。1941 年，C.E.Winn 将之提高到 35。而直到 1968 年，别克霍夫数才更新为 40

• **计算机证明**

高速数字计算机的发明，促使更多数学家对“四色问题”的研究。电子计算机问世以后，由于演算速度迅速提高，加之人机对话的出现，大大加快了对四色猜想证明的进程。就在 1976 年 6 月，在美国伊利诺斯大学的两台不同的电子计算机上，用了 1200 个小时，作了 100 亿个判断，结果没有一张地图是需要五色的，最终证明了四色定理，轰动了世界。

这是一百多年来吸引许多数学家与数学爱好者的大事，当两位数学家将他们的研究成果发表的时候，当地的邮局在当天发出的所有邮件上都加盖了“四色足够”的特制邮戳，以庆祝这一难题获得解决。

四色定理是第一个主要由电脑验证成立的著名数学定理。这一证明刚开始并不被所有的数学家接受。1979 年，逻辑哲学和数学哲学家托马斯·蒂莫兹佐（英语：Thomas Tymoczko）在《四色定理及其哲学意义》一文中提出，四色定理与其证明能否称之为“定理”和“证明”，尚有疑问。“证明”的定义也需要进行再次审视。蒂莫兹佐的理由包括两点：**一方面，计算机辅助下的证明无法由人力进行核查审阅，因为人无法重复计算机的所有运算步骤；另一方面，计算机辅助的证明无法形成逻辑上正则化的表述，因为其中的机器部分依赖于现实经验的反馈，无法转换为抽象的逻辑过程**。即便在数学界中，对四色定理证明的误解也存在着。有的数学家认为证明是杰出的进展，也有人认为依赖计算机给出的证明很难令人满意。也有人认为，计算机辅助证明数学定理不过是对人的能力进行延伸的结果，因为电子计算机不过是依照人的逻辑来进行每一步的操作，实际上只是将人能够完成的工作用更短的时间来完成。还有人将计算机辅助证明和传统证明的差别比喻为借助天文望远镜发现新星和用肉眼发现新星的区别。

> “如果四色问题有一个不依赖计算机的证明，我会更加开心，但我也愿意接受阿佩尔和哈肯的证明——谁叫我们别无选择呢？
> （I would be much happier with a computer-free proof of the four color problem, but I am willing to accept the Appel–Haken proof – beggars cannot be choosers.）”
>
> ***保罗·埃尔德什，1991 年***

针对证明过程冗长、难以理解的问题，哈肯等人也着手对证明进行改良。简化证明的一个方向是寻找更小的不可避免集和更加容易验证的可约构形。哈肯等人很快将不可避免构形集的大小从 1936 个改进到 1476 个。1994 年，罗宾·托马斯等人又将其改进到只包含 633 个构形、32 个放电规则的放电过程推出的不可避免构形集。由于著名的前车之鉴，数学家们对证明进行详细审视，发现了大量缺漏和错误。特别是厄里奇·史密德等人曾经检查人工证明部分的 40%，并发现放电过程中的一个关键性错误。幸好，这些缺陷和错误都是能够修正的。不过，修正的工作也持续了若干年，才最终完成。修正过程中也出现各种传言，说四色定理的证明其实是错误的。1986 年，哈肯和阿佩尔应《数学情报（英语：Mathematical Intelligencer）》杂

志的邀请写了一篇短文，用清晰易懂的语言总结他们的证明工作。1989 年，最终的定稿以单行本的形式出版，超过 400 页。

对于机器证明的可靠性问题，2004 年 9 月，数学家乔治•龚提尔（英语：Georges Gonthier）使用证明验证程序 Coq 来对当时交由计算机运算的算法程序进行形式上的可靠性验证。证明验证程序是一个由法国开发的软件，能够从逻辑上验证一段电脑程序是否正常运行，并且是否达到了它应该达到的逻辑目的。验证表明，四色定理的机器验证程序确实有效地验证所有构形的可约性，完成了证明中的要求。至此，除了机器硬件、软件可能存在问题外，四色定理的理论部分和计算机证明算法部分都得到验证。（**本文末有相关论文链接**）

严格叙述

•拓扑学阐述

最初的染色问题是用几何学的概念描述的。严谨的版本则需要用到拓扑学的概念来定义。设有一欧几里得平面或其一部分，将其划分为互不重叠的区域的集合。一个“地图”为以下划分方式

①将平面划分为有限个区域，使得任意两个区域的交集是空集，所有的区域的并集是整个平面；

②所有区域中，只有一个区域是无界区域，其余区域都是有界区域。

所谓有界区域，是指能够用一个长和宽都有限的矩形覆盖的区域。无界区域则是不能用这样的矩形覆盖的区域。每个区域相当于通俗说法中的“国家”，而区域之间的边界（“国家”之间的“国界线”）则定义为连续不自交的曲线，也称为连续简单曲线。连续简单曲线是指一个从[0, 1]映射到平面$\mathbf{R}^2$的连续函数 c 的像集：$\mathrm{C} = \{\mathrm{C}(\mathrm{t}); \mathrm{t} \epsilon [0,1]\}$ ， 并且要满足：

任意

$$0 \leq \mathrm{t}_1 < t_2 \leq 1$$

，只要不是

$$\mathrm{t}_1 = 0, t_2 = 1$$

，就必定有$\mathrm{c}(\mathrm{t}_1) \neq \mathrm{c}(\mathrm{t}_2)$， 这样说明曲线不与自身相交（没有“打结”的地方）。如果$\mathrm{c}(0) \neq \mathrm{c}(1)$，就称曲线为弧，否则称曲线为**圈**。可以看出，用边界定义地图更为本质：

平面$\boldsymbol{R}^2$中的一张地图是指有限个简单曲线的集合：

$\mathrm{L} = \{\mathrm{C}_1, \mathrm{C}_2 \ldots \mathrm{C}_m\}, m \in N, m \geq 2$，其中

$$\forall 1 \leq \mathrm{i} \leq \mathrm{m}, \mathrm{C}_i = \{\mathrm{c}_i(\mathrm{t}); \mathrm{t} \in [0,1]\}$$

，c_i 为$[0, 1]$映射到$\mathbf{R}^2$的连续函数。并且任意

$1 \leq \mathrm{i} < \mathrm{j} \leq \mathrm{m}$，曲线$\mathrm{C}_i$和$\mathrm{C}_j$要么没有交点（交集为空集），要么交点是两线的一个公共顶点$\mathrm{E}_{i,j} = c_i(\epsilon_1) = c_j(\epsilon_2)$，$\epsilon_1, \epsilon_2 \in \{0,1\}$。

L中每一条连续简单曲线称为地图的**边**。任意边的端点称为**顶点**。可以说，一张地图实际上是由一个简单有界平面图定义的。定义地图的边和顶点后，设所有属于边或顶点的点为**中性点**，其集合设为$\mathrm{N}_L = \{x; x \in C_i, 1 \leq i \leq m\}$，则 L将其余的点划分为若干个道路连通的开集。用拓扑学的语言来说，每个“国家”是$\mathrm{R}^2 - N_L$的一个极大连通子集。或者说，取一个非中性点x，所有能够从x，经过一条不含中性点的弧到达的点构成的集合，就是一个国家。这样定义的

国家必然满足之前所说的特性，只有一个无界国家。要注意的是这里定义的国家必然是没有飞地的。

最后可以定义染色。假设将使用到的颜色编号为1,2,3 …，n号颜色，为地图染色是指一个将地图中的国家映射到{1,2,3, … , n}上的函数。一个可行的 n-染色方案是指使得相邻的国家对应的颜色不同的函数。四色定理说明：每个地图都存在可行的 4-染色方案。

- **图论阐述**

拓扑学版本的四色问题阐述可以转化为更为抽象的图论版本。这里的转化指的是一种对偶的概念。即将一个地图转化为图论中的一个无向平面图。具体来说，是将地图中的每一个国家用其内部的一个点代表，作为一个顶点。如果两个国家相邻，就在两个顶点之间连一条线。这样得到的图必然是一个平面图（不会有两条边相交），而与每个国家选取的代表点无关。四色定理可以叙述为：必然可以用四种颜色给平面图的顶点染色，使得相连的顶点颜色不同。

要注意的是，并非所有的地图都可以转化为图论中的平面图。如果一个国家有飞地的话，就不能用只一个点来代表一个国家。另外，如果一个国家是“国中国”，那么即便可以地图其转化为平面图，也会造成讨论上的不便。但是，“国中国”的着色十分容易解决，因为它只有一个邻国，只需将它染成和邻国不一样的颜色就可以。所以在大部分有关四色问题的讨论中可以忽略“国中国”的情形。同样地，只有两个邻国的情形也可以被忽略。如果规定不能够有四个或者以上的国家有公共边界，那么地图转化成的平面图里面，每个区域都是至多由三条边围成的。这样的地图被称为正规地图。如果任何一个顶点都连出三条边，那么就称其为“三度图”（trivalent map）。可以证明，如果存在四色定理的反例，那么国家数最少的反例必定是三度图。因此在四色问题的证明过程中，常常会假设地图对应的图是三度图

*文章部分资料引用自维基百科

扫描二维码下载 Georges Gonthier 在计算机软件 Coq 的帮助下找到四色定理形式证明（pdf 格式）