

安德烈·塞迈雷迪的工作

W. T. Gowers/文 史永堂/译

1. 引言

安德烈·塞迈雷迪是数学中组合数学领域的杰出人物，特别是在极值组合论这一子领域做出了重要的贡献。稍后我将解释这些术语的意思，但是这里我们首先来介绍一下关于他非凡数学成就的一些枯燥的事实。他的最著名的成就是，在1975年给出了众所周知的埃尔德什（Paul Erdős）和Turán提出的一个具有几十年历史的猜想的证明，现在被称为是Szemerédi定理。这个定理不仅是二十世纪数学的重要贡献之一，而且也是当前大量研究的核心所在。他给出的Szemerédi正则引理，是源于Szemerédi定理证明的一个结果，但是也已经逐渐成为极值组合论的一个重要工具。除此之外，他已经发表了200余篇论文，其中很多论文都描述了重要的进展。我将会选择其中一两个问题来讲解，但是我们应该知道的是，它们仅仅是对很多领域的数学思想有深远影响的巨大成就中的一个小的例子。

那么，什么是组合数学呢？一个可能的定义是组合数学是来研究离散结构的。那么什么是离散结构呢？“离散”这个词是跟“连续”对应的：一个结构是连续的，如果你可以从一个部分平滑地移动到另一部分；如果你不得不跳跃，就是离散的。例如，如果你在建立流体流动的模型，那么你研究的数学结构将是连续的，因为你将考虑类似于不同点的速度和压力的一些东西，并且这些是平滑地变化的。相反地，如果你在建立计算机内部东西的模型，那么你将会对0-1序列感兴趣，这是离散结构的一个例子，因为要从一个这样的序列得到另一个序列，你将不得不从0跳到1至少一次，反之亦然。

另一个离散结构，也许是组合数学中的一个最重要的结构，是图。它包含一些点，以及连接某些点的线。点称为顶点，线称为边。

你可能会认为图是连续的，因为你可以沿着它的边连续地移动。然而，连续的仅仅是图片，而不是图本身。对一个图，我们所关心的是，顶点中的哪些对是由边连接的，这可以被看作为一个简单的列表。例如，如果一个图由一个正方形的顶点和边构成，我们可以称它的顶点为 a, b, c, d ，边记为 ab, bc, cd, da 。

2. Szemerédi 定理

并不是离散结构的所有研究都被划为组合数学。大部分题目的另一个特征是这些问题都可以以容易理解的方式叙述，至少要比许多其他领域的问题要容易得多。同样，证明经常是初等的，不是像歇洛克·福尔摩斯那样用到整个世界，而是仅仅用到某种数学。当一个数学家描述某个证明是初等的时候，这意味着证明不会用到深层次的概念，或者不依赖于之前建立的某些复杂的结果。但是我们不应该认为这意味着证明是简单的：人们可以以极其复杂的方式将一些初等的元素放在一起，而且有些“初等”的证明在数学中是困难的。相反地，一些很高级的证明，当你花足够的时间去理解它们所依赖的定理时，实际上可能是非常简单的。（当然，也确实存在一些简单的初等证明以及困难的高级证明。）

Szemerédi定理是我刚才所讲内容的一个完美的解释。它有一个吸引人的叙述，并且塞迈雷迪给出的证明既是初等的又是极其困难的。让我以解释定理的内容开始讲起。

为了讲解定理的内容，我需要等差级数（或称为算术级数）的概念。一个等差级数是指一个数的序列，每一项都增加相同的大小（称为公差）。所以序列3, 9, 15, 21, 27, 33, 39是一个等差级数，它的公差为6，而序列4, 7, 11, 14, 17, 21不是等差级数，因为相邻两项的差是不相同的（有些是3，有些是4）。

一种理解Szemerédi定理的方法是去想象下面的单玩家的比赛。告诉你一个比较小的数字，比如是5，和一个大的数字，比如是10000。你的任务是在1和10000之间尽可能地选择尽量多的整数，你要遵守的规则是你所选的整数中不



阿贝尔委员会关于塞迈雷迪获奖的海报

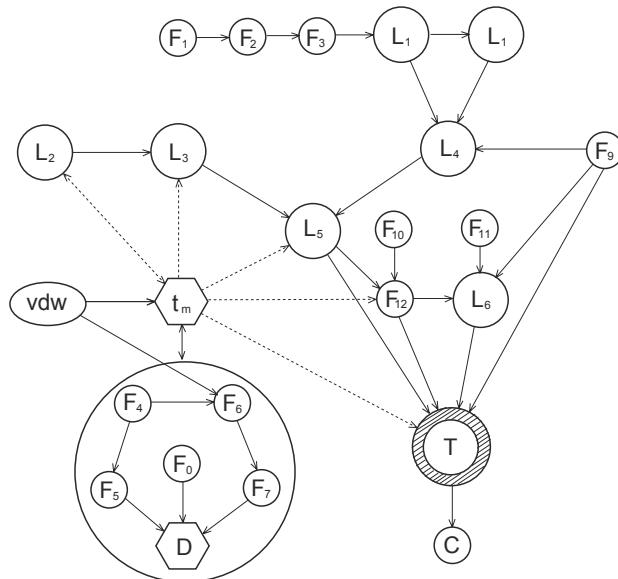
应该包含 5- 项的等差级数。举例来说，如果你碰巧选择了数字 101, 1103, 2105, 3107 和 4109 (以及一些其他的数字)，那么你将输掉比赛，因为这五个数字构成了一个五项的等差级数，他们的公差是 1002。

显然地，最终你是注定要输掉这个比赛的，因为我们可以给出一个既初等又极其简单的证明，如果你要保持足够长的选择，最终你将选择 1 到 10000 中的所有数字，这将包含许多 5- 项的等差级数。但是 Szemerédi 定理告诉我们一些更有趣的事情：即使你使用最好的回避等差级数的策略，在你远远还没接近选择所有数字之前，你就已经输掉比赛了。

为了更准确地叙述这个结果，我将需要一点代数的知识，我的意思是希望用两个字母来表示我们开始时的那两个数字。用 k 表示我们希望避免的级数的长度， n 表示我们希望从中选择的初始数字的个数。(在我们上面的讨论中， k 是 5, n 是 10000。) 现在用 $S(k, n)$ 表示在避免 k - 项等差级数出现的前提下，所能选择数字的最大数目。塞迈雷迪证明了：当 n 很大时， $S(k, n)$ 是 n 的一个非常小的百分数。那么有多小呢？要多小有多小，只要假设 n 是足够大的。

例如，如果你尝试去回避长为 23 的级数，Szemerédi 定理告诉我们存在某个 n (可能非常巨大，但是它是存在的) 满足下面的条件：如果我们用 n 个数字来玩比赛，在我们输掉比赛之前，我们不会选择多于 $n/1000$ 个数字，即总数的 0.1%。而且，这个结果对任何其它的级数长度和任何其它的正的百分数都是成立的。

那么怎么去证明呢？这里我不打算去告诉你，我将再次说明从技巧意义上来说那是初等的，并且我将从塞迈雷迪的原始文章中复制一个图表来尝试使你相信这也是困难的。



3. 为什么我们要关心等差级数的寻找呢？

我在现实生活中并没有看到如下情形的重要性：一个小小的整数集合包含一个长为 10 的等差级数。即使出现这种情形，Szemerédi 定理告诉你你所需要的 n 将会远远大于宇宙中原子的数目，或者是这个数字的指数。这将使定理远离任何实际应用的领域。那么，数学家们到底发现了 Szemerédi 定理的哪些方面是如此的迷人呢？

对这个问题，有很多回答。最显而易见的是，Szemerédi 定理叙述的简单与其证明（以及所有随后发现的证明）的困难之间的对比。通常情况下，一个简单的、自然的数学陈述要么有一个简单的证明，要么有一个简单的反例。然而，有些时候人们惊奇地看到：一个形式无误的问题的解决要远比人们的期望困难得多。这些问题中的大部分被证明太难了，以至于没有人相信在与当前完全不同的新思想出现之前它们将能够被解决 ($e+\pi$ 是否是无理数的问题就是一个例子)。但是有些问题不是这样的：问题的提出很简单并且难于回答，但是它们跟某些问题有足够的联系，使得人们感觉尝试解决它们并不是完全没有希望的事情。Erdős-Turán 猜想就可以归于这一类。

第二个回答是，Szemerédi 定理有很多数学的应用，尽管它没有什么实际的应用。其中非常著名的是 Ben Green 和陶哲轩的一个定理：你可以发现仅包含素数的任意长的等差级数。这个结果并不是直接源于 Szemerédi 定理，但是 Green 和陶哲轩发现了一个极其巧妙的方法将他们的问题转化为一个可以用 Szemerédi 定理来解决的形式。

然而，如果你坚持要求实际应用的话，那么一切并没有丢失，只要你能够接受非直接的应用而不是直接的应用。数学的一个重要的但不值得充分赞赏的方面是，你要证明的结果往往没有你证明它所用的方法更有意思。这一趋势在组合数学中尤为明显，组合数学中的一些公开问题往往越来越受关注，不是因为我们渴望知道它们的答案，而是因为它们包含了许多更一般的困难，这些困难让我们感觉到我们又真正地回到了数学。当一个这样的问题得到解决时，它的解常常包含新的数学工具的发展，这些新的工具能够在许多其他情况下得到使用。

Szemerédi 定理就是这种现象的一个奇妙的解释，正如我们将要看到的。

4. Szemerédi 正则引理

到目前为止我还没有讲什么是极值组合论，那么现在让我开始讲。这里有一个极值图论的问题：如果一个图有 n 个顶点，那么在保证任意三个点都不构成三角形的情况下，它可以有多少条边呢？一般来说，极值组合论中的问题是在



Ben Green(左) 和陶哲轩著名的数论工作用到了 Szemerédi 定理

限制某些其他情况发生的前提下，问某个量可以达到多大。Szemerédi 定理本身就是另外一个这样的例子，它描述了一个这样的问题：在限制长为 k 的等差级数出现的情况下，你可以从 1 到 n 中选择多少个数字。

这些问题自然地分为两部分。一部分是去寻找能够避免你希望避免的结构的例子，以这样的方式使得你感兴趣的量能够尽可能的大。另一部分是去证明如果问题中的量达到某个数值，那么你将不能避免你希望避免的结构。埃尔德什的一个深刻的观察，在很多情况下是解决第一部分问题的一个出奇得好的方法，即去随机地选择你的结构。例如，如果你的结构是 n 个顶点的图，如果每对顶点 x 和 y 是否有边相连仅仅是靠投掷硬币来决定（如果你希望你的随机图包含很少的边，那么你可以使用一枚扁的硬币），那么对某些问题，你将得到非常好的解答。尽管看起来你可能对随机定义的结构不能说什么，但是如果你正在寻找完全确定的事情，那么这差不多是正确的。然而，如果我们仅仅要求接近确定的事情，那么我们就有很多可以说的了，这正是我们要讲的。如果说我们可以一个随机选取的结构几乎确定地具有我们希望的性质，那么我们就可以给出一个弱得多的结论：至少存在一个结构具有我们希望的性质，这样就足够了。

埃尔德什的观察诞生了随机图论，现在已经成为组合数学的一个主要的子领域。作为一个结果，组合数学家们不认为随机图是讨厌的、混乱的东西，而认为从许多方面来说都是容易理解的东西。为了更清楚地表达，我不再说随机图论是一个容易的领域：如果你问关于一个随机图的足够详细的问题，那么回答它们需要非常精密的、困难的概率估计。但是随机图在以下方面具有重要意义，随机图（几乎总）是可预言的并且效果良好。

由于这一背景，Szemerédi 正则引理应该得到大家理解。在介绍正则引理之前，我们假设图被认为是没有结构的东西。最后，当你介绍一个图时，对每对顶点 x 和 y 你需要决定是否用一条边连接它们，并且你的决定没有任何限制。然

而塞迈雷迪意识到，一旦你丢掉你对随机性的恐惧，你就就可以对完全任意的图给出一个有用的结构描述。

在这里我不能给出这个结果的一个精确的叙述，但是大致的思想是这样的。给定任意一个图，存在一种如下的方式将它的所有顶点划分为一些集合，这些集合的数目比较小：如果你取出连接其中任意两个集合的边，那么它们看起来像是被随机选择的（事实上，尽管这个大致的思想是过于简单的，但是对这些目的它将起作用）。简言之，每个图都是由数目很少的随机图构成的。

这告诉我们，可以使用很小数目的数据来给我们的图一个很好的刻画：将顶点划分成引理告诉我们的一些集合，我们只需要粗略地说每对集合之间有多少条边，并且我们知道这些边是以一种看起来随机的方式分配的。这并不能确切地告诉我们这个图是什么样子的，但是更多情形下两个随机图的差别是不重要的，因为它们都是随机的，所以它们都具有你希望一个随机图所具有的性质。

Szemerédi 正则引理很快成为，并且现在仍然是极值图论的一个核心的工具，它的间接的影响，如随后的许多变形和推广，仍然非常广泛。

我许诺过要讨论正则引理的一些间接的实际应用。所以让我以一个具有明显实际重要性的人工智能问题开始，然后回到正则引理。没有人可以否认：如果有人可以设计一个能够从经验中学习的计算机程序，那么这个程序将有无数的实际应用。试图去做这件事的人工智能的一个分支称为机器学习。一个被称为 PAC 学习的著名的抽象模型是由 Leslie Valiant 提出的，是研究机器学习的一种好的结构。[字母 PAC 代表“probably approximately correct”(概率近似正确)]。这引出了一个被归入一般性能测试的纯数学的问题。粗略地讲，你有一个结构，你希望去证明它要么有某个性质，要么它可能非常类似于一个不具有这个性质的结构。例如，给你一个图，让你去证明它包含一个三角形，或者它跟一个不包含三角形的图只有略微的不同。从机器学习的角度来看，我们感兴趣的是这些通常可以被格外快地处理：程序会执行很小数目的简单测试，然后形成一个可靠的假设（也就是，一个概率近似正确的假设）。允许我们去证明这件事的工具是什么呢？那就是 Szemerédi 正则引理。

5. 排序

塞迈雷迪也因成为 Miklos Ajtai, Janos Komlos 和 Endre Szemerédi (简称 AKS) 三人之一而著名。我将论述他们工作的两个最著名的部分，但是我必须强调这只是一个小样本。当我说“最著名”的时候，我的意思是类似于高速公路上的光线：它们非常的昂贵，但是它们也非常的多。

计算机科学中的一个古老的话题是排序算法。给你一

个对象的集合，并且它们中的任意两个都是可以比较的，你的任务是用尽可能少的比较来对它们进行排序。为了研究这个问题，你可以想象这是岩石的一个集合，你要按照它们的重量进行排序，做这件事你所需要的是一副天平，它能够告诉你任意给定的两块岩石哪块更重。你每使用一次天平需要支付一美元，并且你希望花费尽可能少的钱。

一个漂亮的论证说明如果你有 n 个对象，你需要进行的比较的数目至少是 $n!$ 的（以 2 为底的）对数（感叹号表示阶乘，所以 $n!$ 表示 $1 \times 2 \times \dots \times n$ ）。这个论证足够简单以至于我可以在这里给出。 n 个对象的所有可能的排序数目是 $n!$ 。每次你询问的时候，只有两个可能的答案，所以对目前而言你所收到的答案构成的排序的数目，你不能希望将它减少超过 2 的因子。（这有点像给出 20 个问题：每个问题将世界的剩余东西一分为二，如果你不够幸运的话，你得到的答案将是比较大的那一半，所以一般来说，你不会比将可能性的集合减少 50% 做得更好。）因此， k 步之后，你不能将可能序列的集合减少多于 2^k 的因子。于是在最坏的情况下，步骤的数目必然为满足的 2^k 至少为 $n!$ 的某个 k ，否则的话，一定存在多于一个由你的信息所构成的序列。这等价地说 k 一定至少为 $n!$ 的对数。

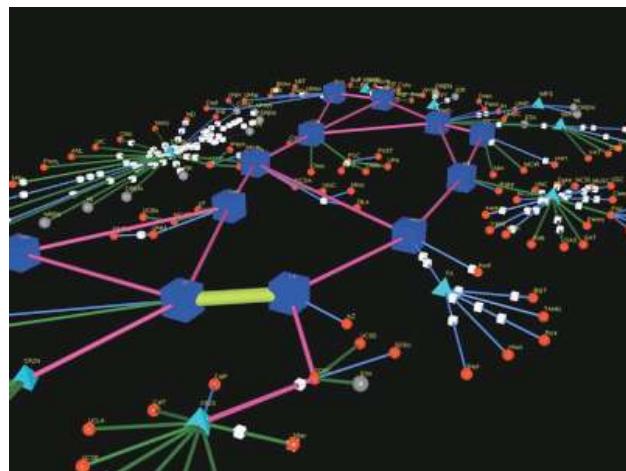
$n!$ 的对数大概是 $n \log(n)$ （即 n 的对数的 n 倍）。有趣的是，目前已知的排序算法是通过近似比较的数目而得到的。也就是说，为给出之前的答案，即要求许多接近理论最小值的比较，存在已知的方法用来决定对哪两个对象进行比较。例如，冯·诺依曼在 1945 年发明了一种实现这个问题的方法，称为归并排序（Merge Sort）。

在 1945 年底，塞迈雷迪仅仅五岁，因此，排序算法这一课题早在塞迈雷迪研究数学之前就已经完全出现了。

让我们回到我们希望按照重量排序的岩石问题，并且将这一游戏做轻微的变化。这一次你将拥有尽可能多的天平，这样你就可以同时进行多次比较。所以你要做的是组织一系列的称重，每次称重中你可以对尽可能多对的岩石称重。一个明显的限制是同一块岩石不能放在不同的天平上，所以在每次称重中每块岩石至多被比较一次。你现在的目标是使用最少的称重次数对岩石进行排序。

因为我们已经知道大概需要 $n \log(n)$ 次比较，并且每次称重你最多可以做 $n/2$ 次比较，所以显然称重的总数将会大约为 $2 \log(n)$ 。在很长一段时间里，人们能否成功地得到这一理论最小数值是一个公开问题，这就是 AKS 解决的问题。他们提出了一个极其聪明的排序方法，使得称重次数确实达到 $\log(n)$ 。用计算机科学的语言来说，AKS 发现的是用于排序的一种快速并行算法。

这里我不能完全的描述他们的方法，但是可以给出这一方法的某些想法。假设你以 1000 块岩石开始，并且将它



图论既困难又有实际应用

们分成两组，每组 500 块。（在这一步，你绝对没有任何比较两块岩石重量的思路。）你要做的事情是将一组中的岩石跟另一组中的岩石进行配对，同时做 500 次比较，然后将较轻的那些岩石放到标记为 L 的一组，较重的那些放到标记为 H 的一组。如果 L 中的所有岩石都比 H 中的所有岩石要轻，那么我们将非常高兴，但是目前为止，我们绝对没有任何理由来相信这是正确的。无论如何，我们将重复这一步骤。当然，为了得到更多的信息，我们现在希望以一种不同的方式对岩石进行配对。

根据埃尔德什的方法，我们可能认为将 L 中的岩石和 H 中的岩石随机地配对并重复这一步骤是一个好的方法，的确如此。也就是说，每一次如果 L 中的一块岩石比我们从 H 中捡出的岩石要重的话，我们就将它们交换一下，如果轻的话，我们保持不动。

如果我们保持将 L 中的岩石和 H 中的岩石随机地配对，那么存在一个一般的趋势，最终 L 中的岩石较轻一点， H 中的岩石较重一点。而且随着程序的继续，这个趋势越来越显著：如果 L 中的大部分岩石已经是较轻的了，那么它们将不太可能被移到 H 中， H 中的岩石也不太可能被移动到 L 中。AKS 证明在常数次比较之后（即一个不随 n 的增大而增大的数字）， L 中的绝大多数岩石都比 H 中的绝大多数岩石要轻。

如果 L 中的所有岩石都比 H 中的所有岩石都轻，那么我只需在每一组中重复这一程序，并且大约 $\log(n)$ 步之后我们就可以结束了。然而，我们所知道的是这对几乎所有数目的岩石是正确的，所以尽管 AKS 确实达到了这一结论，但是他们要完全证明这个问题还需要付出很大努力。

关于这个算法的最后一点：我上面所描述的算法是一个随机算法，因为所有的对都是随机对。然而，存在一类非常有用被称为扩展图的图，它可作为随机图的替代来使

用。这里，你可以用扩展图的边来决定如何去给岩石配对，而不是靠投掷硬币的办法，并且这时算法仍然是起作用的。这就是去随机化（derandomization）的一个例子，这是理论计算机科学的一个基本思想，它意味着算法可以以一种完全确定的方式运行。

6. Ramsey 数 $R(3, k)$

你所期望的图中的一个三角形是：三个顶点彼此相连。一个独立集是一个顶点的集合，他们彼此没有边相连。这里给出一个简单的论述来证明，一个具有 k^2 个顶点的图一定包含一个三角形或者包含一个大小为 k 的独立集。等价地说，如果一个图既不包含三角形，也不包含大小为 k 的独立集，那么它的顶点个数一定小于 k^2 ，这实际上是我要证明的。

论述如下。令 v 是图中的任意一个顶点。与 v 关联的任何两个顶点都不会有边相连，因为那样的话我们就有三个顶点。因此，与 v 关联的所有顶点构成一个独立集。于是 v 不能跟多于 $k-1$ 个其他顶点关联（因为我们假设这个图不包含大小为 k 的独立集）。

现在我们设想来尝试寻找一个大的独立集。我们做这件事的一种方法是去寻找一个顶点的序列 v_1, v_2, \dots ，确保我们选择的每个新的顶点跟之前选过的任何顶点都不关联。我们能做到这点吗？假设目前为止我们已经选择了顶点 v_1, v_2, \dots, v_j ，其中每一个都关联于至多 $k-1$ 顶点，所以它们之间至多关联 $(k-1)j$ 个顶点。因此，当选择一个新顶点时，我们必须回避至多 $j + (k-1)j = kj$ 个顶点。所以只要顶点的数目超过 kj ，我们就可以扩展这个序列。如果我们想得到从 j 到 k ，那么我们需要的顶点数目要超过 $k(k-1)$ 。因为 $k^2 > k(k-1)$ ，所以如果我们有 k^2 个顶点，我们将能继续我们的序列直到至少 v_k ，于是这给了我们一个大小为 k 的独立集。

我们真的需要那么多个顶点来保证一个三角形或者一个大小为 k 的独立集吗？我们可以给出一个比 k^2 稍微强一点的结果： $k(k-1)+1$ 个顶点就足够了。但是存在 $k(k-1)$ 个顶点的图，它既不包含三角形，也不包含大小为 k 的独立集吗？如果不存在，这样的图可以有多少个顶点呢？满足这些的最大可能的数称为 Ramsey 数 $R(3, k)$ 。

上面的讨论（根据数学研究的标准）是非常简单的，但是确定是否可以改进它被证明是非常困难的。这就是 AKS 解决的问题：他们证明 $R(3, k)$ 至多为 $k^2/\log(k)$ 乘以一个常数因子。如果你真的希望欣赏一下这个成就，你就应该花一点时间自己来考虑一下这个问题，但是如果你没有时间，那么你至少要记住这个问题已经有几十年的历史了，而且 15 年后 Jeong Han Kim 在另一篇著名的文章中证明了这个结果是最好的：即 $R(3, k)$ 不会超过 $k^2/\log(k)$ ，而且实际上等于 $k^2/\log(k)$ （仅相差一个常数因子）。

为了证明他们的结果，AKS 可以比上面讲到的简单论述做得更好。极其粗略地，他们用的是同样的讨论，但是用一种更谨慎的方式处理的。回忆一下基本的想法是选择顶点 v_1, v_2, \dots ，并且回避你所选顶点的邻点。我们可以描述这个程序如下。从整个图出发，我们选择一个顶点 v_1 ，并且扔掉它的所有邻点；然后从图的剩余部分选择顶点 v_2 ，并且扔掉它的所有邻点。我们这样继续做下去。因为任何顶点的邻点个数都不会超过 $k-1$ ，所以每一步我们不会扔掉太多的顶点，因此我们可以持续相当长的一段时间。

现在如果我们在程序的任何一步中，发现了一个顶点，它的邻点数目远小于 $k-1$ ，那么我们将会非常高兴：我们将选取那个顶点，并且我们不需要扔掉图的那么多个顶点，所以程序可以持续的更长。但是我们可以找到这样的顶点吗？最开始的时候，没有任何原因。但是我们确实对这个程序给出了某个控制：当我们选择一个顶点 v_i 的时候，我们扔掉它的邻点。如果这些邻点本身有很多的邻点，那么当我们将它们扔掉的时候，我们也将扔掉很大数目的边。所以基本的想法是以这样的方式选择序列 v_1, v_2, \dots ：尽量减少没有被扔掉的图的那部分边的数目。AKS 证明将这个想法继续下去是可能的，并且最终会延长这一进程使得算法将会持续大约 $\log(k)$ 次，只要那个简单的论断成立。

上面我所暗示的讨论并不是 AKS 给出的原始证明，但是描述起来比较容易。他们早期的讨论也是非常重要的，因为它是称为半随机方法的一个较早的例子，这里我将不再讨论，但是我想说这是另一种具有很多应用的技巧。

7. 结论

一些数学家因为一个或者两个定理而著名，另一些数学家因为巨大的、重要的一些一流结果而著名。偶尔，也会有数学家因为两者而著名。没有 Szemerédi 定理和 Szemerédi 正则引理的讨论，塞迈雷迪的工作将不会完整。然而，对塞迈雷迪来说，他的工作远远多于这两个定理，他已经发表了 200 余篇论文，正如我在开始时提到的，71 岁的他并没有任何慢下来的迹象。他应该得到阿贝尔奖，这是极其恰当的。即使我仅仅了解他工作的一些表面情况，我还是希望通过这个小样本描述，至少能够让大家认识到推荐他获奖的一些理由。

作者简介：W. T. Gowers 是英国皇家学会院士，剑桥大学数学教授，因其将泛函分析与组合数学联系起来的贡献于 1998 年获菲尔兹奖。