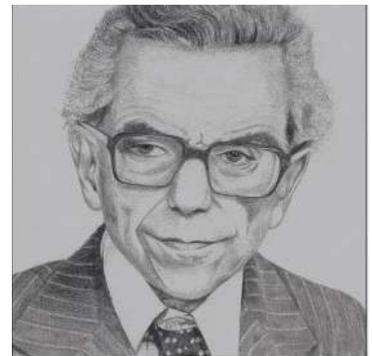


你的埃尔德什数有几？

—— 数字背后的人际网络图

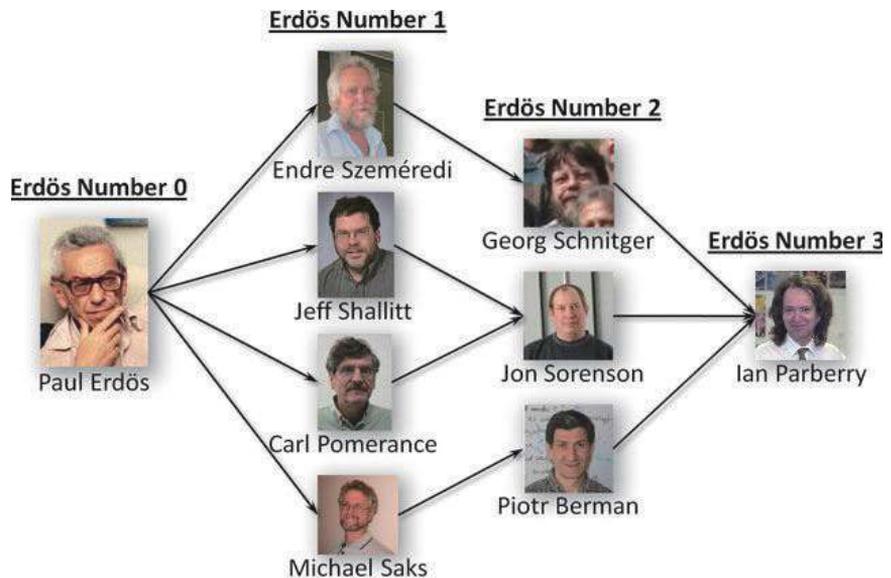
Jerrold Grossman / 文 赵京 / 译

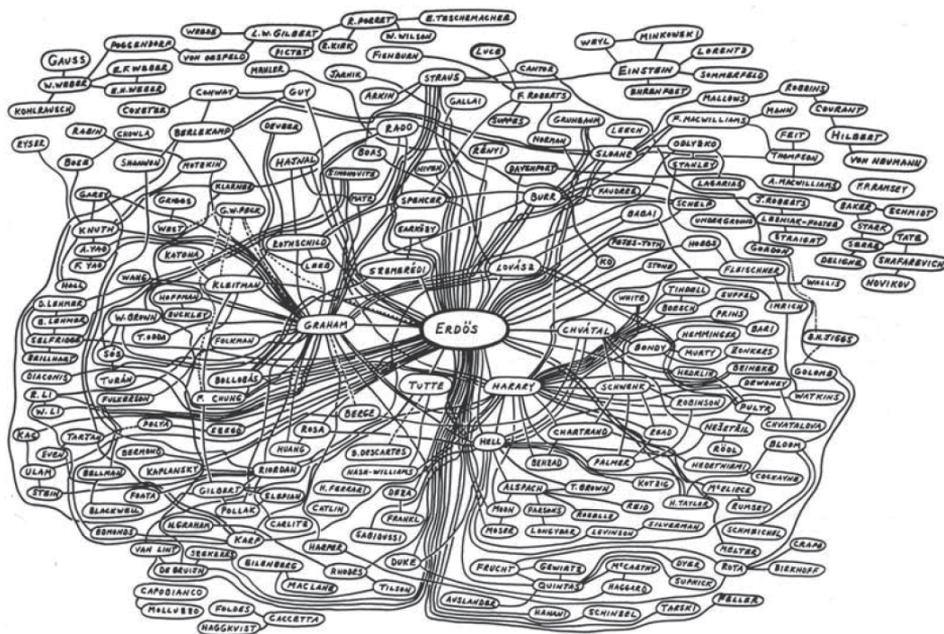
早在脸书 (Facebook) 和“凯文·培根六度空间”游戏 (The Six Degree of Kevin Bacon) 诞生之前，数学家们就从保罗·埃尔德什 (Paul Erdős) 现象开始懂得多种人际和社会网络之说了。事实上，现在大多数的社会网络研究的框架都可以追溯到埃尔德什 40 年代关于随机图的研究，特别是他和匈牙利数学家瑞尼 (Alfréd Rényi) 合写的那篇独创性的文章《随机图的演化》¹。另一方面，我们要感谢埃尔德什生前和约五百名合作者共同发表的庞大数目的数学文章，大约 1500 篇。这惊人的数字刺激了数学家们在他们的合作活动中寻找相连性，并逐步将它们升华成数学问题来探讨。数学家戈夫曼 (Casper Goffman) 1969 年发表在《美国数学月刊》上的一篇文章²正式定义了数学家们早已使用的专用词汇——埃尔德什数 (埃数) (Erdős Number)。严格的定义是这样的：保罗·埃尔德什本人的埃数是 0；他的每一位合作者的埃数是 1；埃数 1 的合作者，但又没有与埃尔德什合作过的人，拥有埃数 2，并以此类推。如果找不到某人和埃尔德什任何一个层次的合作者有关系的话，那么此人的埃数是无穷。



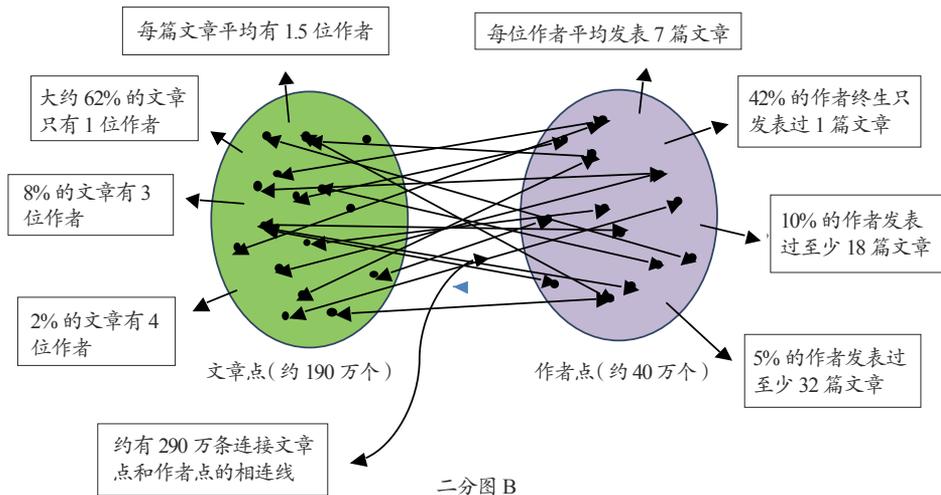
埃尔德什画像

美国数学会出版的《数学评论》(Mathematical Reviews) 杂志在互联网上提供了自 1940 年起正式发表过的所有数学文章的准确数据库 (以下简称“MR 数据库”，译者





注)。通过 MR 数据库里的信息（本文数据是基于 2004 年的一项研究，我们感谢美国数学会提供此项数据），我们不难发现由埃尔德什合作衍生出的有趣现象，下面我们就用数学语言来描述一下。假设让字母 B 代表一个二分图（Bipartite Graph），图上一部分的点（Vertices）代表 MR 数据库里的所有文章（文章点）；图上另一部分的点代表 MR 数据库里的所有作者（作者点）；每一篇文章和它的每一位作者之间用一条线（Edge）相连（相连线）。那么，这样画出的 B 有大约 190 万个文章点；40 万个作者点；290 万条相连线。通过 B 可以算出 MR 数据库里每位作者平均发表 7 篇文章，标准差是 18。大学终身职位评估委员会可能更愿意了解作者发表文章数的中位值，它是 2。事实上，MR 数据库里 42% 以上的作者只发表过一篇文章；40% 的作者发表过 3 篇以上的文章；30% 的作者发表过至少 4 篇文章；20% 的作者发表过至少 8 篇文章；10% 的作者至少发表过 18 篇文章；5% 的作者至少发表过 32 篇文章。现在，让我们从另一个角度来看这组数据。二分图 B 显示每一篇发表的文章平均有 1.5 位作者；大约 62% 的文章只



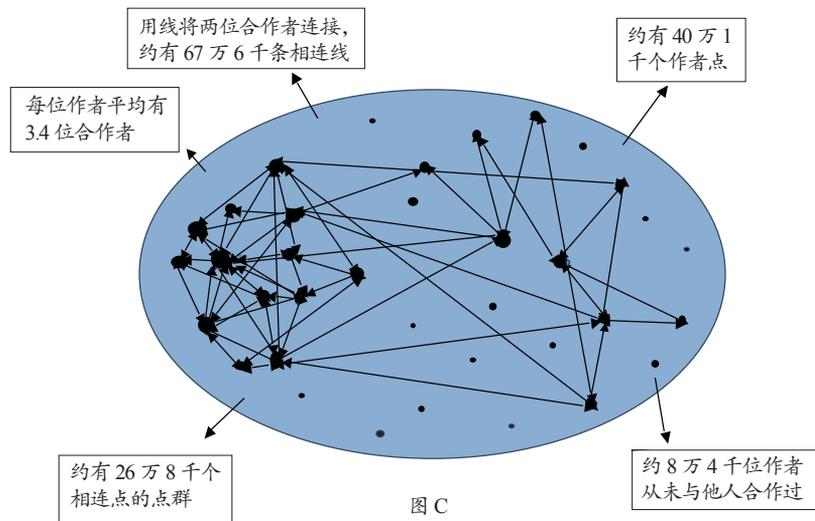
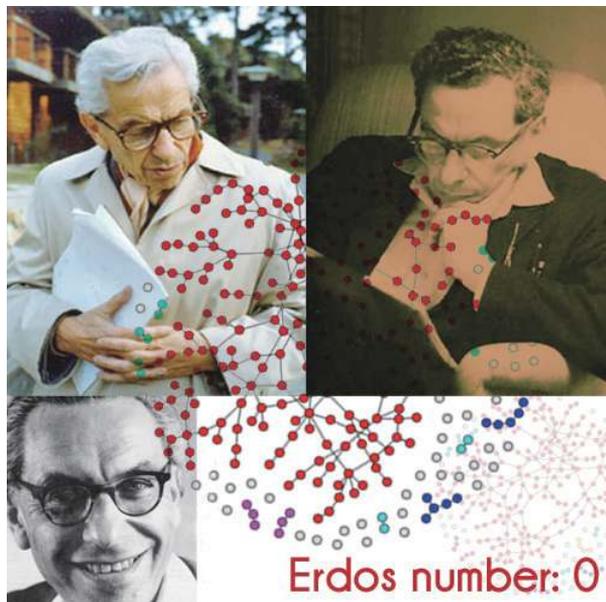


图 C

有 1 位作者；27% 的文章有 2 位作者；8% 的文章有 3 位作者；只有 2% 的文章有 4 位以上的作者。只有 1 位作者的文章比例从 1940 年的 90% 逐步下降到了现在的不足 50%。

通过二分图 B 我们可以画出一个新的图 C 来更精确地描述数学界的合作关系。图 C 包含 40.1 万个点来代表 MR 数据库里的作者。如果其中两位作者曾合作发表过文章的话，就用一条线将两人连接起来。这样画出的图 C 有大约 67 万 6 千条线，也就是说每位作者平均有 3.4 位合作者，标准差是 6.6；每位作者拥有的合作者人数的中位值是 1。社会网络结构的最新研究指出，上述合作者的数目应该遵循幂律。因此，拥有 x 位合作者的数学家的人数应该和 x 的某次方成正比，次方数应该是在 -2 或者 -3 左右。的确，当我们用这个模型模拟我们的数据时，所得到的次方数大约是 -2.97。图 C 里有一个含有大约 26 万 8 千个点的群，群里的点是相连的。在社会网络中，这种内部小范围的相连性是很常见的。图 C 里剩余的 13 万 3 千个点中，有 8 万 4 千个点是孤立的，剩余点里最大的相连点群仅含有 32 个点。换句话说，MR 数据库里有 8 万 4 千位作者从未与其他人合作过。



那些想知道自己埃数的数学家们当然想了解上述特大相连点群里每个点距埃尔德什点的距离，最长的距离是 13。如果我们用罗格斯大学 (Rutgers University) 的盖尔范德 (Izail Gelfand) 教授或者普林斯顿大学的西奈 (Yakov Sinai) 教授做为群的中心点的话，群的半径是 12。盖尔范德和西奈教授的埃数是 3。特大相连点群里两点之间的平均距离是 7.7，标准差是 1.2；距离的中位值是 7；四分位值分别是 6 和 8。如果我们把 MR 数据库里 75% 的作者都考虑进