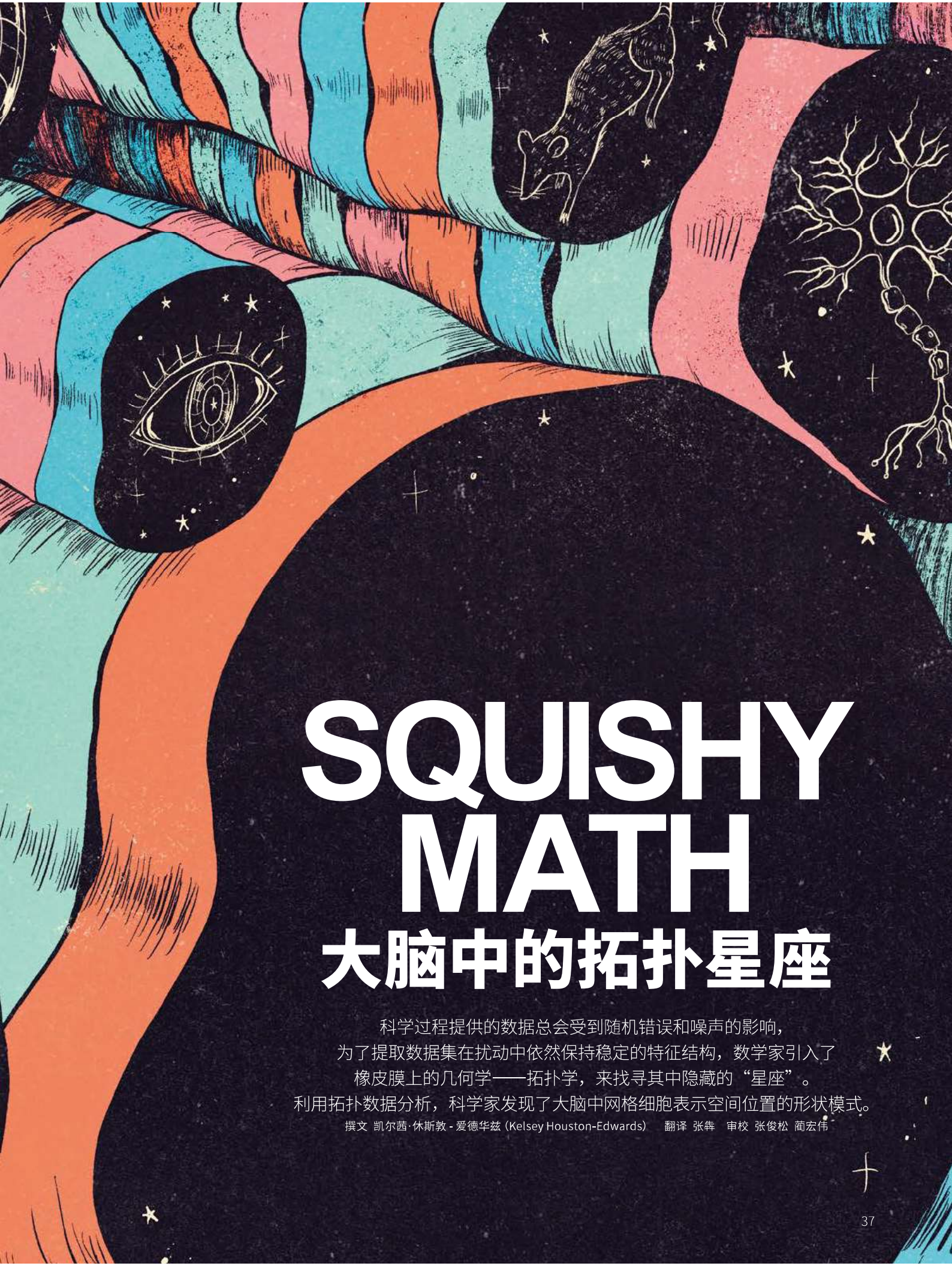
图片来源：Evangeline Gallagher



SQUISHY MATH

大脑中的拓扑星座

科学过程提供的数据总会受到随机错误和噪声的影响，为了提取数据集在扰动中依然保持稳定的特征结构，数学家引入了橡皮膜上的几何学——拓扑学，来找寻其中隐藏的“星座”。利用拓扑数据分析，科学家发现了大脑中网格细胞表示空间位置的形状模式。

撰文 凯尔茜·休斯敦 - 爱德华兹 (Kelsey Houston-Edwards)　翻译 张犇　审校 张俊松 蔺宏伟

凯尔茜·休斯敦 - 爱德华兹是一位数学家，也是一位科普作家。她曾参与线上节目 PBS 无限系列（PBS Infinite Series）的创作，并担任过该节目的主持。

挪威科技大学的数据科学家本杰明·阿德里克·邓恩（Benjamin Adric Dunn）向我展示了一张图片，图中散布着间隔不一的点，它们如同巨石阵中的石头一样粗略地排列着。对于人类而言，点阵排列形成的图案看上去一目了然。“当我们扫视这张图片的时候，这些点很明显排布成一个圆，”他说道。然而，对于计算机而言，却很难通过图像识别算法辨认出这个简单的形状。“因为它总是忽视全局。”

许多科学过程都涉及循环或重复的环状模式，但实际上计算机很难发现这种模式。对于想从大量数据点中识别出环状模式的科学家而言，这是一个严峻的问题。我们常常想象数据就像夜空中的星星一样，是一个个悬浮在数学空间中的点。一个数据点可能用来表示一个地理位置，比如船只在公海的位置就能在经纬度构成的二维空间中标记为一个点。同样地，基因也可以被绘制在数学空间中，只是由于基因复杂的属性，这个空间的维度有时会多达上百个。通过这种方式，两个具有相似 DNA 序列的基因就能在这个空间中表示为两个临近的点。数据中环状模式的意义通常取决于应用背景，比如，船只位置数据中的圆环可能代表船只迷航了，而基因数据中的圆环则可能展现了一种演化关系。

通常，这些数据点集的空间维数都很高，也极其复杂，我们无法直接用肉眼分析和研究。所以，科学家需要一个计算机能够理解的精确指令集来检测数据中的圆环。然而，许多标准的数据分析技术都是基于研究直线和平面的数学分支——线性代数。为了检测环，研究人员转而求助于拓扑数据分析（Topological Data Analysis，TDA），它提供了一种截然不同的视角。

与线性代数研究中相对简单的刚性结构不同，TDA 的基础是拓扑学，这是一门研究可任意拉伸扭曲形状的数学分支。拓扑学领域的研究者假设所有形状都可以任意弯曲，因此拓扑学也被称为橡皮膜上的几何学。

就像根据形状能将散落的星星划分为一个个星座一样，TDA 能帮助数学家从散乱的数据点集中提炼出清晰的形状。数学家首先选择一些数据点作为虚拟支架的节点，以此构建可以扩展到上百维的复杂结构。这样生成的结构不仅囊括原始数据的大部分要素，还能以更明确的形式将它们呈现出来。他们从拓扑的角度研究这些结构，寻找那些在支架拉伸、弯曲变形时也能保持不变的特征。

拓扑学在数据分析中的用处就在于它能揭示数据集中定性而非定量的性质。科学研究中，实际测量的过程总伴随着随机错误和噪声，它们常常会使数据发生扰动。但拓扑学能帮助识别出数据结构中不受噪声影响、保持稳定的部分，从而揭示系统的稳健性。“现实世界提供的数据在

精彩速览

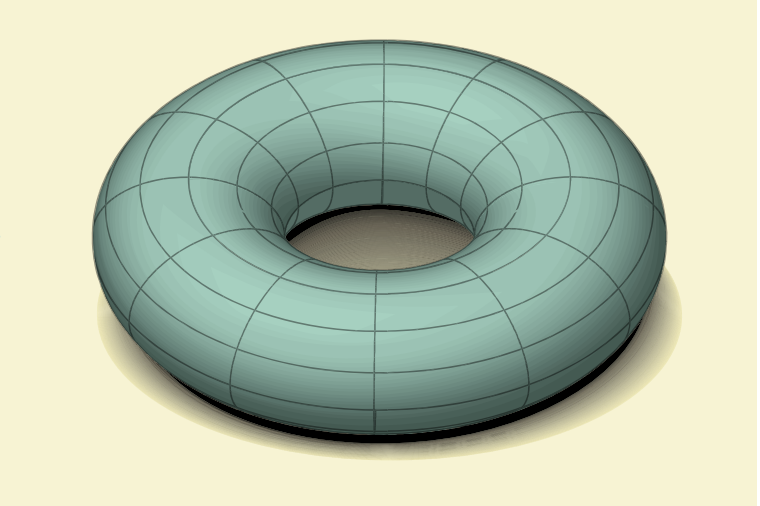
计算机总容易忽视全局，这导致它很难发现大量数据中隐藏的环状模式。研究可任意拉伸形状的拓扑学，能帮助识别数据中稳定的结构特征。

拓扑学家很善于对形状中的洞进行分类和计数，通过分析脑神经数据转换的图所构建的一系列单纯复形，能提取出包含重要信息的“顽强洞”的列表。

过去的实验发现，大鼠激活脑中网格细胞时所处位置呈规律且重复的点图案，借助拓扑数据分析，科学家发现环面可能是网格细胞表示空间的固有形式。

很多情况下都是“黏糊糊”的，”美国宾夕法尼亚大学的数学家罗伯特·格瑞斯特（Robert Ghrist）说：“所以你也必须使用黏连的数学来应对。”

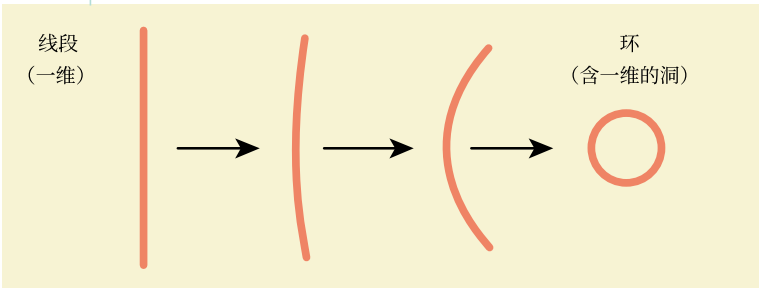
从日常的生物过程到药物分子结构，不同的科学过程都可能提取出大规模数据。而数学家和科学家正在共同努力，希望从中找出不同寻常的拓扑形状。这其中，最令人感兴趣的也许是有关大脑结构的研究。数学家开始用拓扑学来探索神经元是如何跨脑区相互作用，从而对不同环境和刺激做出反应。最近，邓恩在与神经科学家的合作中发现，特定的脑细胞使用环面（描述甜甜圈外表面的数学名称）来映射周围的环境。

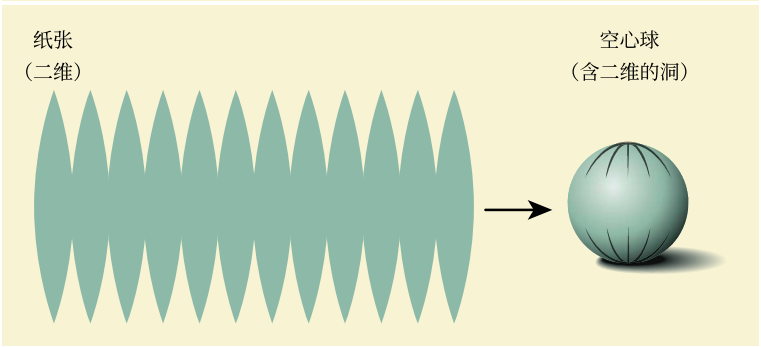


甜甜圈与咖啡杯

虽然拓扑学家通过扭曲橡胶膜来进行研究，但他们会非常谨慎地保留橡皮膜上洞的数量。这意味着，他们从不会打一个新洞或挤压封闭一个已经存在的洞。有一个经典的数学笑话讲道，拓扑学家分不清甜甜圈和咖啡杯：因为它们都有且只有一个洞。

通常，拓扑学家根据空间维度对洞进行分类。一个由一维线段两端粘接而成的闭环，比如数字 0，我们称它包含一个一维的洞。而对于二维平面，比如一张纸，如果把它的边缘粘起来会得到一个像空心球一样的物体，因此我们称这个空心球包含一个二维的洞。





制图：珍·克里斯琴森（Jen Christiansen）

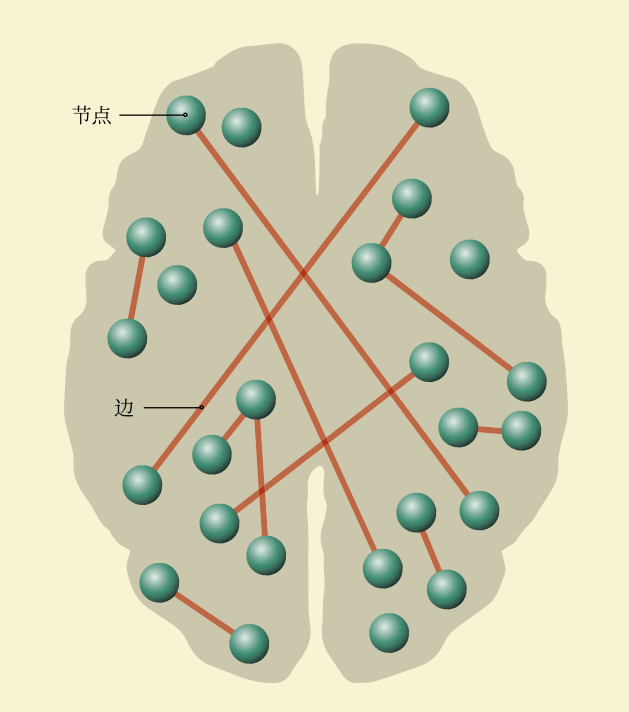
更高维的形状可以包含更高维的洞。例如，以一维洞和二维洞类推，三维洞是通过“封闭”一个三维空间（比如一个立方体）形成的。但捏合三维空间构成洞的过程只能从更高维度，比如四维视角下观察到，这显然超出大多数人（可能是所有人）的能力范围。

此外，一些形状可能同时包含多个不同维度的洞，比如有一种充气球，孩子可以握住把手坐在上面弹跳。这种球内部的空心部分是一个二维的洞，而实心把手则形成一个一维的洞。如果只是考虑三维形状，这看上去很简单，但若想扩展到超出我们认知的高维形状和洞，就需要引入数学辅助。拓扑学中有许多精确的方法能对更高维形状中的洞进行计数，这种能力也被证明对研究大脑神经元的活动很有帮助。

美国印第安纳大学（Indiana University）的神经科学家奥拉夫·斯波恩斯（Olaf Sporns）将大脑看作一个大规模的交通网络。它的道路和基础设施是由神经元以及连接神经元的突触构成的。大脑活动时发出的电信号和化学信号就沿着这些道路传输。“受到神经元以及突触这些真实的信号通路限制，我们没法从全局观察这些传输模式的动态变化，”斯波恩斯说道。而现实中，大脑的信号传输模式会一直随着我们的行动和思考动态地改变。

当我们的目光扫过大脑图谱时，它看上去像一个点的集合，这些点代表神经元。其中一些点被线条连接起来，这表示这些神经元之间存在突触。数学家将大脑图谱这种抽象结构称

为图：一个由边连接的节点集合。图有效地降低了大脑的



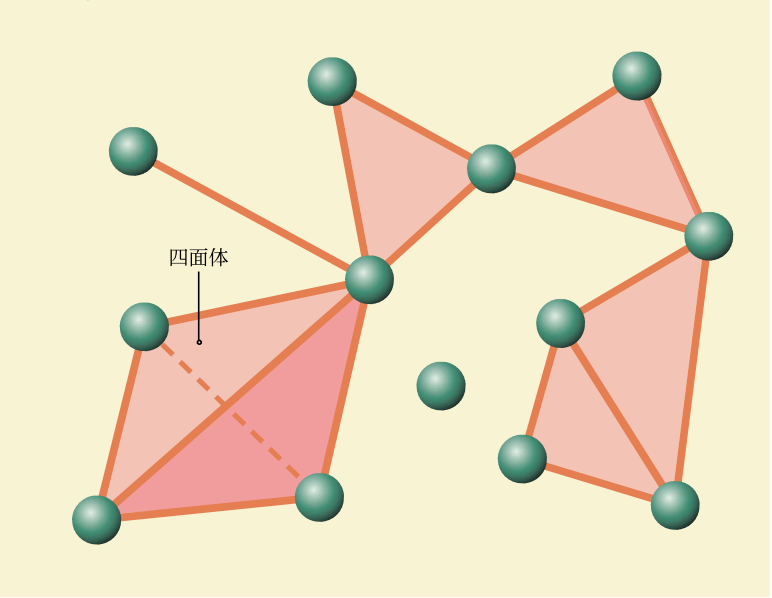
生物复杂度，同时却保留了环路的整体形状。在构建数学模型时，这种在简便性、可分析性与实用性之间进行的权衡是很常见的。

神经元连接图非常像一张复杂的网：数量庞大的神经元紧密地交织在一起。为了研究这张复杂的网，2017年，瑞士联邦理工学院的数学家凯瑟琳·赫斯（Kathryn Hess）完成了一项在当时令人震惊的工作：使神经元连接图变得更加复杂。她分析了“蓝脑计划”（Blue Brain Project）的数据，并以此构建了一个模拟啮齿动物新皮层（参与大脑中部分高阶认知功能的执行）活动的大规模计算机仿真模型。模型不仅模拟了所有独立的神经元，每个神经元还能通过突触连接到其他的仿真神经元。而这些仿真神经元间如何连接、以及它们何时触发动作电位都由生物学原理和实验室啮齿动物的实验数据来决定。

仿真实验可以显示大脑的传输模式——神经元通过放电来响应刺激。然而，与观察真实大脑的鸟瞰图不同的是，模型可以被按下暂停键，从而为科学家提供定格画面，显示出哪些突触在受到给定刺激会发生放电反应。同时，由于定格画面显示了数据点和它们之间的连接，比如如果连接两个神经元的突触正在放电，那么很显然它们之间形成了突触连接，因此这种定格图片非常容易被转换成图。根据这一现象，赫斯构造了数学家称为单纯复形（simplicial complex）的拓扑对象，它能将简单的图转化为一个庞大的形状。

一个单纯复形由多个不同维度的三角形搭建而成。例如，在蓝脑计划数据转化的图中，通过三个突触连接并能互相传递信号的三个神经元，就构成了一个空心三角形。为了将这个结构扩展成一个单纯复形，数学家用一个实心的二维三角面填充这个空心三角形。以此类推，可以用更高维的三角形相似物来填充有更多连接的神经元集群。例如，拥有四面三角形界面的实心三维锥体——四面体就可以填充一个同时激发的四神经元群。

赫斯和其他科学家观察到，能同时发生受激放电反应的神经元数量最多为八个，因此这个单纯复形中最大的部分是一个七维三角形。可以想象，许多元素叠加，会导致形成一个具有多重维度的对象：一个三角形可能从一个四面体中延伸出来，并在一点上与另一个三角形相交。此外，



数学家和科学家还模拟了轻触啮齿动物胡须触发的大脑传输模式，并由此得到了一系列定格画面。他们将每张定格画面对应的图转换成一个单纯复形，并使用拓扑工具分析其形状如何随时间发生变化。

在接受刺激后，单纯复形会像一个巨大的乐高积木模型一样快速生长，不断根据所受刺激添加越来越高维的部分。整个对象最高可能达到三维或四维，然后就会飞速地消失。“这些越来越复杂的结构由刺激产生，直到它突然溃散，”赫斯说道。

对于拓扑学家而言，三条线段连接形成的三角形与一个空心圆是一样的，因为一种形状总可以弯曲成另一种形状。由于赫斯和同事以啮齿动物大脑仿真模型构建的单纯复形是七维的，所以它们包含的洞最高也能达到七维。他们的分析表明，随着形状的生长，洞的数量也在增加。在积木搭建的最高点，这个结构可能包含了数量惊人的二维和三维洞——比随机或根据其他生物过程构建的单纯复形所能产生的洞要多得多。这种特殊的洞的发展模式揭示了神经元反应的高度组织性；而洞的复杂性可能也标示了大脑思考过程的一个基本特征。

不会消失的洞

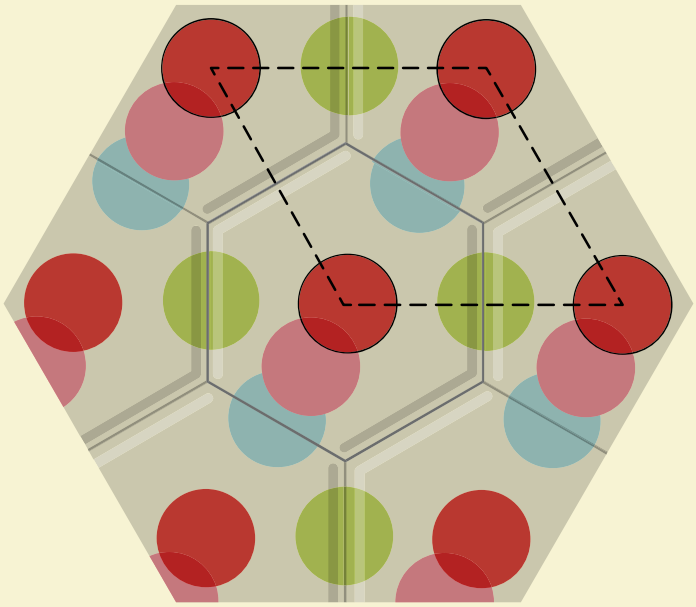
更多情况下，数据被表示为抽象数学空间中漂浮的孤立点，而数据点间并不存在有预设的明显连接。要想应用 TDA，数学家需要搞清楚如何将这些点连接起来。然而，将星星连接成星座的方式是多种多样的。为了找到这些潜藏的图形，数学家使用了一种被称为“持续同调”（persistent homology）的技术。拓扑学家通过分析在不同尺度下构建的一系列单纯复形，能够提取出数据云的本质特征。

为了建立第一个单纯复形，他们撒出尽可能宽的网，将每个点与其余所有点连接，从而形成一个密集的网络。用实心体填充这个网络会产生一个几乎没有可识别特征的高维单纯复形。但是数学家还需要将这个复形与其他在更小尺度上连接数据形成的复形进行比较。因此，接下来，他们在同一数据云上撒下更窄的网，只连接相邻的点，从而得到一张更稀疏的网络，并用它来构建第二个单纯复形。由于第二张网包含的数据点更少，所以它对应的单纯复形也会由更低维的形状组成。研究人员会用一系列更小的网重复这一过程。美国俄亥俄州立大学的数学家兰东尼·埃德蒙兹（Ranthony Edmonds）说道：“通过这些操作，你会得到这个复杂结构在每一个尺度下的不同快照。”

可以说，每个单纯复型都是从同一组离散数据点中，根据不同尺度抽象提取出的星座图。拓扑学家研究了这些迥异的形状，特别是记录下不同形状在各个维度的洞的数量。他们对在多种不同尺度下持续存在的洞特别感兴趣。一些洞短暂地出现然后就会消失，但有些顽强的洞——在一系列不同尺度下都保留下来的洞——则指向了数据最本质的特征。因此，TDA 可以将复杂、乱糟糟的数据简化为一个关于“顽强洞”的简单列表，就像 JPEG 格式文件压缩图像的方式一样。“这是一种将数据提炼出真正重要信息的方法，我们由此获得了更易处理的信息，”格瑞斯特表示。

有时用这种方法发现的洞有很直接的解释。美国东北大学的数学家何塞·佩雷亚（Jose Perea）和一个计算生物学家团队利用持续同调发现了周期性的生物过程——那些在规律间隔下重复的过程。例如酵母菌的代谢周期或小鼠的生物钟。“循环或重复是什么？”佩雷亚问道。“从几何上来说，就像是在你观察的对象所构成的空间中穿过某种环。”

TDA 还可以帮助研究人员设计新的药物。一些化合物通常是通过调整现有药物的分子结构发现的。但是，即使是对于机器学习算法，分子结构也极为复杂且难以分析。为了设计新的药物，计算机需要简化现有分子的表示。有很多方法可以做到这一点，但美国密歇根州立大学的魏国卫（Guowei Wei）团队选择将分子简化为它们的“拓扑特征”。这是一种基于化学物质拓扑特征的描述——本质上是使用持续同调获得的信息集合，比如各个维度的顽强



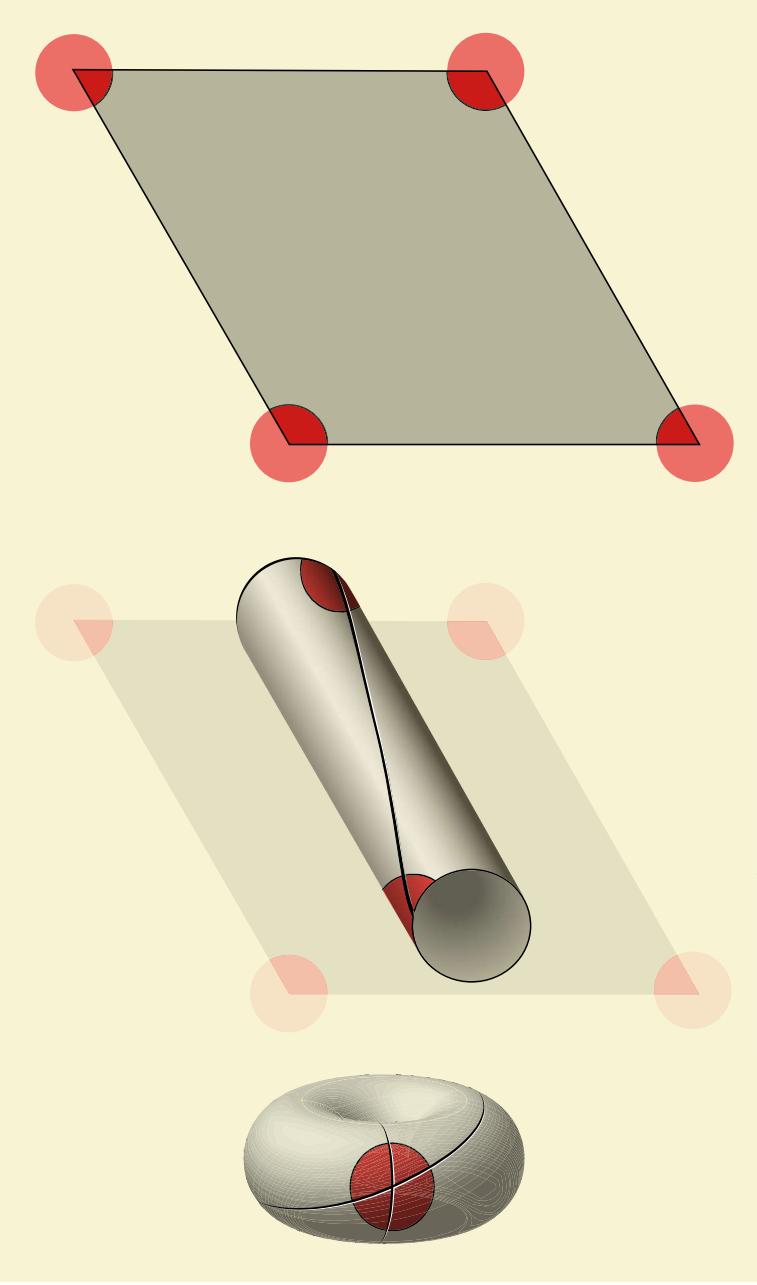
四个不同的网格细胞（红、粉、蓝、绿），
每个网格细胞的激发模式都呈现相同的形状

洞的数量。

脑中的环状模式

TDA 最有趣的应用可能是在大脑组织的最基础层面——一种单一类型的神经元。2014 年，约翰 · 奥基夫（John O’Keefe）和研究搭档梅 - 布里特·莫泽（May-Britt Moser），以及爱德华 · 莫泽（Edvard Moser）因为分别发现位置细胞和网格细胞（当动物处于特定位置时被激活的神经元类型）而获得诺贝尔医学奖。“这两种神经元充当了位置传感器，”美国宾夕法尼亚州立大学的数学家卡丽娜 · 库尔托（Carina Curto）说道。

当大鼠处在环境中的某几个位置时，大脑中对应的网格细胞就会被激发。为了搞清网格细胞和大鼠所处位置之间的关系，神经科学家挑选了一个网格细胞进行研究。每当这个网格细胞被激活，科学家会在模拟地面的计算机模型上绘制一个点来标记此时大鼠所处的位置。当大鼠在方形盒子里自由移动时，模型呈现出一种有规律且重复的点图案，数学家将这种图案描述为六边形格子。格子上的点代表了这个特定网格细胞被激发时大鼠所处的所有位置。他们对多个网格细胞重复这一记录过程，并用不同的颜色标记对应的点图案。结果显示，每个网格细胞对应的点整体呈现相同的几何形状模式，但相互之间存在偏移，像是杂乱的瓷砖一样覆盖整个盒子（见第 41 页右下图）。

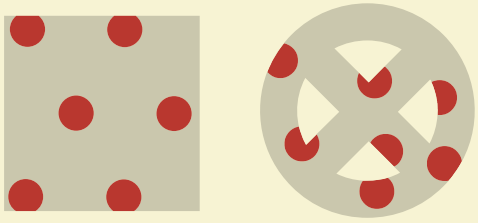


神经科学家想确切地知道网格细胞是如何映射空间位置的。而本质上，他们是在寻找产生六边形图案的模板。例如，想象一个印着各种卡通人物的圆形橡皮图章，在纸面上滚动图章，它会留下一条线。每隔一段距离，米老鼠的形象就会出现在这条线上。但所有这些图像都来自这个橡皮图章上的同一个地方。理解滚动图章能生成图案很容易，但反过来，这个问题就变得极具挑战性：如何根据产生的图案生成对应的模板图章？

在表示网格细胞激发时大鼠所处位置的彩色点平铺图中，四个相邻的红点平铺在平行四边形的四个角。就像图章中重复出现的米老鼠图像一样，相同颜色的所有红点都对应着同一个单独的网格细胞。所以拓扑学家识别出所有的红点，并用一种称为“粘接”的操作将平行四边形折叠成甜甜圈的形状。首先，他们把平行四边形的两个对边粘接在一起，形成一个有两个红点的柱面：一个在上面，一个在下面。而后，他们把圆柱体弯曲，将两端粘接在一起形成一个环面。这样，平行四边形的四个红色的角就变成了甜甜圈上的一个点。而其他每种颜色的一个点都刚好会出现在这个环面上。因此，如同圆形图章是一行卡通人物的模板一样，环面则是网格细胞如何表示方盒底部空间位置的正确模板。

当大鼠在盒中跑来跑去时，神经科学家可

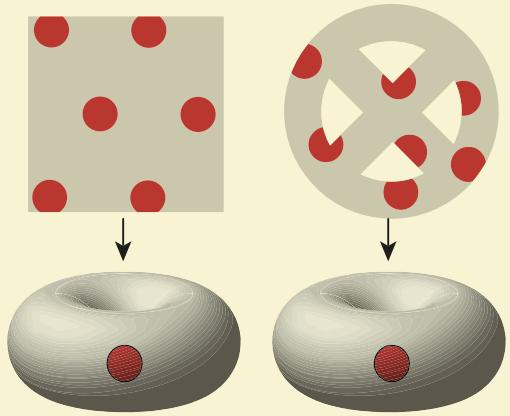
以很轻松地观察到这种模式。但当大鼠在其他测试场地（比如带有辐条和轮毂的自行车轮形状的场地）移动时，就很难想象这种模式了。每个网格细胞仍然在多个位置激发，



当大鼠在方盒区域内运动时，
网格细胞激发模式的简化视图呈现为规律排布的结构（左），
当大鼠围绕一个轮状圆形区域运动时，
网格细胞激发模式的简化视图并没有呈现规律排布（右）。

但科学家却无法确定其实际的分布图。因为这些点的排布并没有呈现出明显结构。

在 2022 年 2 月《自然》（*Nature*）刊登的一篇论文中，一个由数学家和神经科学家组成的研究团队（也包括邓恩）使用网格细胞来检验一种称为连续吸引子网络的理论，该理论预测某些神经元以特定的模式连接在一起——即使动



网格细胞在盒状和轮状场地的激发模式都可以映射到一个环面上。

物处于不同的环境中，这种模式也不会改变。为了检验连续吸引子网络理论，研究人员需要确定不论大鼠发现自己身处何种环境中，网格细胞激发的图是否总能形成一个环面。也就是说，科学家需要在杂乱的神经活动数据中寻找环面——而这对 TDA 而言是个再恰当不过的任务。

与之前的模型不同，研究人员没有标记盒中单个网格细胞激发时的位置，而是研究了整个网格细胞网络的集体活动。每隔一定时间，他们会用一串 0 和 1 组成的字符串记录网络的状态，而每个字符是 0 或是 1 则取决于对应单个网格细胞是否被激发。从数学家的角度来看，这串长字符串是非常高维的空间中的一个点。事实上，当研究人员记录下系统在不同时刻的状态时，他们也在不断地积累高维数据点。这些点描述了网格细胞的激活模式随时间演变的过程，但这些数据太过复杂以至于无法直接用肉眼观察研究。

在使用了一些标准技术来简化数据后，研究团队在不同尺度下连接数据点生成了一系列单纯复形，并通过分析单纯复形来计算系统的持续同调。与之前一样，当大鼠在方盒范围内奔跑时，数据点形成了一个环面。但真正的测试其实是从绕轮状场地运动的大鼠身上获得数据。令人兴奋的是，这些数据又一次形成了环面。

研究人员甚至能够从一只睡着的——可能是在做梦的——大鼠身上收集数据。他们再次发现了一个环面，这意味着，无论大鼠所处环境或状态如何，这种形状都会持续存在。这一发现支持了连续吸引子网络理论。甜甜圈形状似乎是网格细胞表示空间的固有方式。

拓扑数据分析的许多应用都是基于强大的新型计算工具才得以实现。“如果计算机科学家没有认真地开始构建算法，这一切都不会发生，”英国牛津大学的数学家维迪特·南达（Vidit Nanda）说道，“如果它无效，或者它不能很好地扩展，那么无论这个理论有多漂亮，都没有人愿意使用它。”

多亏了这些技术，拓扑学（直到最近似乎还只是一个抽象而有趣的数学分支）的应用正在蓬勃地发展。“它的应用越来越强大，”美国斯坦福大学的数学家，TDA 的先驱之一贡纳尔·卡尔松（Gunnar Carlsson）表示。“我们真的跨越了鸿沟。”

本文译者　张犇是浙江大学数学科学学院的博士生，研究方向为拓扑数据分析和深度学习。

本文审校　张俊松是厦门大学信息学院人工智能系的副教授，研究方向为脑认知与智能计算和计算机图形学。

蔺宏伟是浙江大学数学科学学院的教授，研究方向为计算机辅助几何设计、计算机辅助拓扑设计和量子图形学。

扩展阅读

The Math of Making Connections. Kelsey Houston-Edwards; April 2021.