

一个表情包如何一夜间火遍全网?一种产品是怎样闪电般地占据了整个市场?一场疾病又为何会突然全球性暴发?这类问题的答案就藏在经典的数学分支——渗流理论中。

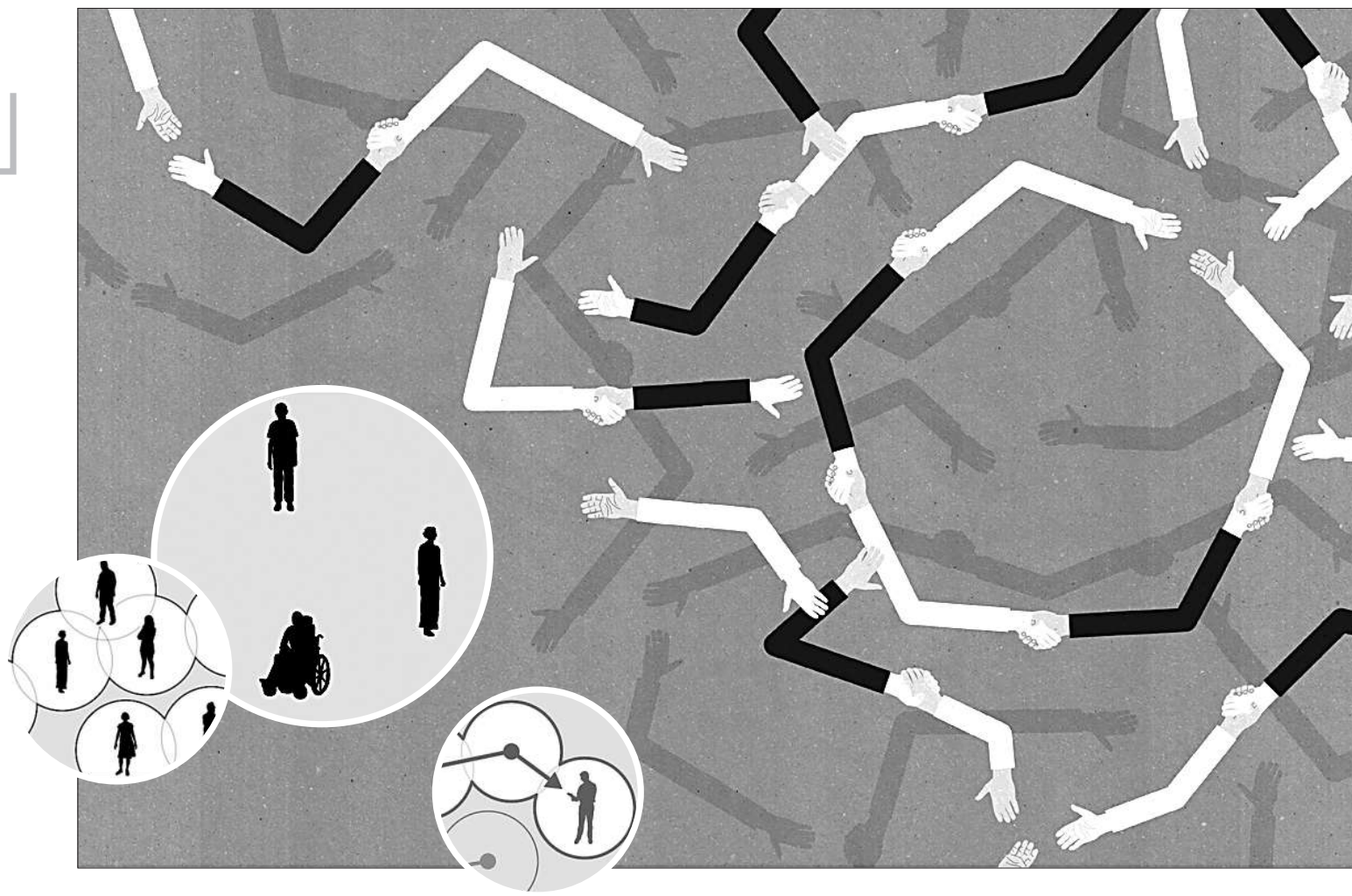
非0即1: 网络中的数学

撰文 凯尔茜·休斯顿-爱德华兹(Kelsey Houston-Edwards)

翻译 蔡格非 冯煜阳 杨 鹏

引言

当你按下“发送”时,你可能以为短信会直接从自己的手机传输到朋友的手机上。事实上,长途传递信息通常需要通过蜂窝网络或互联网完成,这两种网络都依赖于集中调度的基础设施,而这些设施可能会遭到自然灾害等因素的破坏。现在,有一些应用程序正试图帮助用户与临近的手机实现直接的信息传输。这些应用程序能让加密的信息悄悄地从一部手机传到另一部,从而将发信人和收信人连接起来。换言之,这些手机以网格网络(mesh network)或自组织网络(ad hoc network)的连接方式实现了没有中心调度的灵活通信。在这样的网络中,任意两部手机进行通信时,都需要通过短距离内的相互连接来实现。那么,有一个问题出现了:在一个网格网络中,需要同城有多少人相连接,才能保证全城通信?



1 突然的巨变

一门名为渗流理论(percolation theory)的数学分支给了我们出人意料的答案:只需几个人就能让一切变得不同。最早加入这个通信系统的用户会构成若干互不相连的小规模网络。然而一旦用户的密度(单位面积的用户数)超过一个关键的阈值,就会突然出现能够连接远距离用户的信息通道。科学家们将网络连接性的这种快速变化描述为相变(phase transition),这一概念同样也能用来解释冰的融化或水的沸腾等物质状态的突然变化。

渗流理论能研究在这种网络结构中随机创建或删除连接产生的后果。数学家将网络结构设想为边(也就是线)连接顶点所形成的网络——每个顶点代表一个对象,例如电话或人,而边则代表其中两个顶点之间的特定关系。渗流理论可以追溯到20世纪50年代,它的基本观点是:随着网络中连接数量的逐渐增加,将存在无穷多个点彼此间有路径相接,或者用数学的语言来讲,一个无穷大的连通簇(infinite cluster)将会在一瞬间突然出现。

2 牵一发而动全身

1957年,英国数学家西蒙·拉尔夫·布罗德本特和约翰·迈克·哈默斯利首次将流体的滤过过程(例如油渗入多孔岩或者水渗入咖啡粉)抽象成名为渗流的纯数学问题,用一个网络来代表岩层,岩石结构中的孔隙通道为顶点,允许流体在其间流动的道路或裂缝表示为边。我们很容易想象,石油在裂缝较多的岩石中流动得更远。布罗德本特和哈默斯利通过渗流理论预测,在理想化的情形下,当裂缝的密度超过一定的阈值,石油将突然渗透到几乎整个岩层而不再仅限于一些小区域。

地质学家曾使用渗流理论的一个版本来研究有裂缝的岩石中岩团的大小,这些研究与水力压裂法开采石油以及理解地震发生的原理息息相关。为了建立地震模型,地震学家会创建与观测到的裂缝规模和密度相匹配的渗流网络,然后根据调整不同裂缝连接起来的概率来解释应力。随着应力增加,裂缝集群扩大,地震在修复之间不可预测地产生。研究者可以通过调整渗流过程的参数,使缝隙愈合或是再次裂开,以模拟余震或者更长期的变化。

渗流理论还可用于阐明更小规模的物理和化学过程。以聚合过程为例:在这个过程中,单体,也就是小而简单的分子结合在一起,形成更大的集群,称为聚合物。在渗流理论的框架中,每个单体可以作为一个节点,两个相邻的单体可能会自发地形成化学键,对应着网络中的一条边。如果它们连接的可能性不断增加,系统最终会达到渗流阈值,形成远比单体分子庞大的聚合物。这也是明胶溶于水能凝固成果冻的原因。

岩石裂缝或构成聚合物的网络极其复杂。尽管我们几乎不可能精确描述它们的结构,但布罗德本特和哈默斯利表示可以将其近似地描述成一些更易于分析的重叠图样。最简单的例子是由一个个正方形组成的格点网

现在,科学家正在致力于回答:这样的相变何时会发生?对于任何一个给定的网络,如何找到相当于冰在0°C融化或是水在100°C沸腾的时刻?一个表情包如何一夜间火遍全网?一种产品是怎样闪电般地占据了整个市场?一场地震在何时降临?一个手机网络如何能在刹那间实现全面连接,或者一种疾病何时开始在全球蔓延?渗流理论能提供解答这类问题的理论基础。

数学家通常倾向于研究具有几何对称性的无限网络,因为这样便于进行理论计算。一般只有无限网络具有一瞬间发生的相变,术语称为骤然相变(sharp phase transition)。现实世界中的网络计算起来则难度较大,这些网络一般规模有限,且常常非常混乱。不过,这些网络也有相变,只是相变较为平滑。在当下的时代,各种复杂的连接使世界的联系更加紧密,例如为人们提供能源的电网,联系着人们的社交媒体,甚至是传染病的传播网络。因此,渗流理论也与我们息息相关。

格,它看起来就像一张无尽的图纸:网络中顶点依次排列,并由四条边连接到邻侧的点。

为了理解液体是如何穿过这个网络的,不妨将网络中的边想象成一条要么开放要么闭合的水管。所有水管都独立地以相同的概率开放或闭合。于是,所有或开或闭的水管将形成一个随机的网络,其中包含一些连通的区域(连通簇),所有的顶点都由一系列开放的水管连接。如果你往这种连通的任何一个节点中注水,那么水流就会通过那些开放的水管流到这个簇中的所有其他节点。

渗流理论关注的就是网络的连通性,即上述的连通簇究竟有多大。但“大”是一个模糊的概念,并不容易得到形式化的数学表述,因此数学家常常用无穷大来替代那些很大的数。那么,我们讨论的核心问题就变成了:是否存在一个无穷大的连通簇?德国魏尔斯特拉斯应用分析和随机过程研究所(WIAS)的数学家贝内迪克特·雅内尔说:“对我们来说,回答‘是否存在这种簇’比回答‘有多少这种簇’,或者‘这些簇有多大’要容易得多。”

事实上,一个无限大的网络中包含一个无穷大的连通簇(以下简称为“无穷簇”)的概率总是0或1,这是由于渗流过程会受到概率论中一个一般性理论的制约。该理论叫作零一律,由苏联数学家安德里·柯尔莫哥洛夫在20世纪30年代提出。

零一律告诉我们,有限的改变无法干扰本质上无限的现象。所以,在无限大的网络中找到一个无穷簇的概率必须处于某个极端的位置——不是0就是1。这一概率不会产生更细微的改变,比如从0.81变为0.82。换句话讲,一个无限大的网络要么有一个无穷簇,要么一定没有,非此即彼。因此,将有限条水管闭合或开放,对是否存在无穷簇没有任何影响。对是否存在无穷簇的概率要么是0,要么是1。那么到底是哪一种呢?

3 找出阈值

答案取决于水管开放的概率。想象你有一个用于控制这个概率的表盘。当表盘的指针转到最左端时,代表概率为0,水管总会闭合。一旦所有的管道都被关闭,灌入某个节点的水不会往任何地方流,这时找到一个无穷簇的概率为0。当你顺时针转动指针,水管开放的概率会增加,开放管道的总数也越来越多。当指针转到最右端时,该概率为1,所有水管都开放,并且灌入某一个节点的水最终会流入所有其他节点,此时找到

4 连通网络

在物理系统中,格点网络(lattice)是很好的渗流模型。以断裂的岩石为例,孔隙的位置是固定的,但它们之间的裂缝是随机形成的。现实世界中其他的网络甚至还要复杂得多,比如前面提到的神经网络,这一网络中顶点的位置会不断变化,而其中的边,或者说连接,只有在两部手机彼此足够接近(蓝牙通信的有效范围约为10米)的情况下才能形成。类似这样的网络中,顶点可以存在于连续空间中的任何地方。因此我们需要用另一种被称作连续渗流(continuum percolation)的模型来描述它们。

像任何数学模型一样,这种网

5 传染病传播网络

我们用一些网络模型模拟石油在岩石中的流动或电话间的通信网络。事实上,这些网络模拟了这些系统的真实空间结构;如果两个顶点所代表的对象是相连的,就在网络中用一条边将它们连通。然而传染病的传播网络更加复杂,这是由于其连接关系是由某种病毒在人与人之间的传播方式决定的。一个人在一夜之间度过一小时的感染者可能在短短几天内就将病毒传向全国甚至整个大洲。

最简单的流行病学模型会将某一区域的人群简单地分为易感人群、感染人群和康复人群三类,但一般忽略了人与人之间的复杂联系结构。在这种模型中,感染者会将传染病随机传播给易感人群中的其他人,并且易感人群的每个人被感染的概率相同,例如宿舍中的学生与城市中的普通居民感染的可能性是一样的。易感人群感染率取决于基本再生数,即每个感染者造成的新感染者的平均数量,我们将它记为 R_0 。显然,如果 $R_0 > 1$,则说明病毒正在传播;反之,如果 $R_0 < 1$,则说明疫情正在消退。

但实际上,人与人之间的相互关系会影响传染病传播的整体趋势。例如,2003年暴发的严重急性呼吸综合征(SARS)最初的 R_0 值在2.2到3.6之间,然而现任得克萨斯大学新冠肺炎建模联盟的负责人劳伦·安塞尔·迈耶斯在2006年的一篇文章中指出,“通过简单的计算不难看出”,SARS的病例数“比预期要

无限簇的概率为1。

如果你将指针缓慢地从关到开转动,管道打开的概率会逐渐增大,看起来找到无限簇的概率也会从0到1逐渐变化。但事实上,这一变化是瞬间发生的。零一律指出,该概率不会取0到1之间的值。对于正方形网格而言,存在无限簇的概率会在指针恰好处于正中央时突变,这时水管开放与闭合的概率相等。表盘上这个关键的位置被称为渗流阈值。不论网络的形状是什么样的——例如三角

网络或三维正方体网格——渗流理论的基本问题仍然相同:阈值在哪里?管道开放的概率需要多大,才会有足够多的开放连接来保证无穷簇的存在?

答案取决于无限网格网络的精确形状,而要找到它并非易事。即使要证明几乎最简单的系统——正方形网格——的阈值是 $1/2$ 也是一项令人望而却步的挑战。正方形网格的这个问题的最终在1980年被数学家哈利·凯斯滕解决。现在,尽管经过了数十年的努力,但我们仅能

从局部到全面的连接。”雅内尔说。对于神经网络应用程序的设计者来说,找到渗流阈值是一个非常实际的工程问题。想要调控这个网络中的旋钮,一种方式是改变设备的功率。神经网络公司go-Tenna的首席科学家拉姆·拉玛纳坦说:“关键问题是,为了保证网络连接,你应该如何设定传输功率?”如果功率和连通程度存在线性关系的话,答案就相当直观了——每一次小幅增加功率都会导致连通程度按一定比例增加。但是,渗透解成水管系统中那个控制表的旋钮,顺时针任意转动一点点都会增加连通程度。在这些模型中,“存在一个开关,你可以‘咔嚓’一下就实现

低得多”。她认为,这种差异正是因为预测的时候会认为“所有易感人群被感染的概率相同”,而这假设忽略了人群关系的复杂性。并且,上述估计的 R_0 值是基于SARS病毒在公寓和医院中的迅速传播得出的。与普通人群相比,这两个地方的人“彼此密切接触的概率异常之高”。但由于SARS感染者通常很快就会出现严重的病情而前往医院,因此他们无法将SARS病毒传播给更多的人。

在传染病的传播网络中,连接顶点的边表示某种具体的传播关系,例如在一个显示艾滋病传播的网络中,两个人如果交换了体液,那么他们所代表的两个顶点之间就有一条边连接。新冠病毒的传播网络则有所不同:如果两个人紧密接触却没有呼吸防护,那么他们之间就有一条边相连。显然,诸如禁止商业活动和限制旅行之类的限制措施也是减少病毒在人群中,即点与点之间传播的方法。找到控制疫情传播的对策,或者说让这个传播网络断开的途径,正是流行病学学家们面临的挑战。

实际的疾病传播网络非常复杂,难以准确描述。而且即使知道网络的确切结构,也很难进行数学分析。此时,计算机模拟和大数据分析可以用于预测未来的病例数,对比例保持1米和2米的社交距离所产生的效果,以及定量分析学

网络或三维正方体网格——渗流理论的基本问题仍然相同:阈值在哪里?管道开放的概率需要多大,才会有足够多的开放连接来保证无穷簇的存在?

答案取决于无限网格网络的精确形状,而要找到它并非易事。即使要证明几乎最简单的系统——正方形网格——的阈值是 $1/2$ 也是一项令人望而却步的挑战。正方形网格的这个问题的最终在1980年被数学家哈利·凯斯滕解决。现在,尽管经过了数十年的努力,但我们仅能

保证网络的连接性。另一种拨动旋钮的方式是控制节点的密度。具有固定范围的网格网络需要一个用户密度的临界值,这种网络最有可能在音乐节、足球比赛等人员拥挤的活动中提供范围较广的连接。纽约布鲁克林的Red Hook等社区正在使用网格网络,他们将永久性节点固定在建筑物的顶部,从而增强互联网接入。网络中许多必要的硬件和路由技术仍在发展中,但这并不妨碍我们对未来应用的一些大胆设想,比如在无须依赖任何额外基础设施的前提下,自动驾驶车辆可以利用网格网络直接进行通信交流,分享交通状况或道路障碍的信息。

一个顶点度数的可能取值以及取这些取值对应的概率。就传染病的传播网络而言,其代表的是一个感染其他人(或被多少人感染)的可能数量以及相应的概率。为了理解度数分布如何影响渗流的阈值,以迈耶斯为代表的数学流行病学学家生成了数千个网络样本,这些样本除了具有相同的度数分布之外是完全任意的。迈耶斯说,通过这种“控制变量”的方法,人们能够看出度数分布在传染病的传播网络中所起到的作用。如果这些生成网络样本的性质与实际网络相匹配,那么这些生成网络中的度数分布或者其他数学特征就很可能与传染病传播有关。如果这种匹配非常完美,“那么你的数学结论就会像你的模拟结果一样”。

研究表明,在网络的度数分布较分散,或者说顶点度数的取值范围较广的情形下,它的渗流阈值会下降。也就是说,假设有两个网络,一个网络中的每个人都有大致相同数量的接触者,另一个网络中的一部分人有很多接触者而另一些人则较为孤立,那么在后面这个网络中,传染病将更容易传播。澳大利亚墨尔本乐卓博大学的数学流行病学学家乔尔·米勒较为直观地解释了这一结论:“如果我的接触者数量是你的10倍,那么我被感染的可能性就是你的10倍,同时我传播疾病的可能性也是你的10倍,因此我在传染病传播网络中的重要性是你的100倍。”

网络中顶点的度数分布是指

6 未来网络

渗流理论常常被用来模拟许多“传染性”的现象。例如一类表情包最开始在社交媒体上缓慢传播,之后却突然出现暴发式扩散。渗流也可以应用于经济领域的模型,描述一种特定的产品是如何伴随着人们在社交中的分享和推荐而快速占领整个市场。受到人际交流影响的投票模型(voter model),也显示出了这样的门槛效应。

与数学家研究的无限的、整齐有序的网络相比,现实样本中得到的网络模型是有限但杂乱无章的。尽管有限网络不会像无限网络那样,从一个一个小的连通分支瞬间转化为一个几乎处处连通的结构,但这种变化仍然非常迅速。为了理解这类过程,研究网络的学者们在数学理论和计算机模拟之间不断来回探索。较简单的网络能引导他们建立精确的计算机模型,而计算机模拟的结果又会告诉他们如何从手绘的简单网络出发,洞察真实世界的本来面目。

许多重要的新冠病毒传播模型整合了来自其他网络的信息。学校的课表、交通路线以及医院员工日程表都能够形成特定的网络,并且每个网络都会影响流行病的传播。加利福尼亚大学戴维斯分校的复杂网络学者雷萨·迪苏扎(Raissa D'Souza)说:“我们生活在一个由相互依赖的网络组成的系统中,我们不能只研究其中的某一个而不考虑网络间的相互影响。”每一个网络都是一个复杂系统,有着自己的突发性行为。研究者正在逐渐将这些网络组合起来,形成一个更加复杂的系统。但目前还没有清晰的理论框架来研究网络相互嵌套的系统。理解这些网络是如何相互影响的,将会是未来的重大挑战。

韦斯皮尼亚尼说:“我们身处的地方,既不是一个个孤立的气泡,也不是一个所有事物都完全混合的地方。我们生活在一个充满接触和交流的世界中,我们关注推特账号,而渗流模型就能出现在这些事件中。”对于这些数学模型取得更深的理解,“可以在未来有所作为。”渗流模型是很容易被接受的,它为数学家提供了新的游乐场,也为科学家提供了实际的应用。但这些不同的模型有一个令人惊讶的统一特征:它们的变化中都有一个枢轴点(pivot point),当系统处于该点附近时,只需少数几个连接就可以将整个系统连为一体。随着世界中的联系变得越来越紧密,理解这些关键变化的必要性也越发迫切。

(作者系美国数学家兼记者,曾为美国在线节目“PBS无穷级数”撰稿并主持该节目)