



观点分裂现象中的定律

(香港科技大学 章琪、韩一龙 编译自 Anna Demming, *Physics World*, 2024, (1): 37)

早在2006年，加拿大医生、自称“先知”的Bill Deagle就曾声称飞机尾迹被故意装入了一种外星的致病的硅基生命。他警告说：“这是一种像蜜蜂或蚂蚁一样具有智慧并且会反击的硅基生命体。”尽管没有证据表明硅基生命的存在，Deagle的说法还是在社交媒体上不断出现。2023年3月，脸书上的一篇相关帖子获得了超过37000个赞和33000次分享，以及大量的赞同评论。

在社交媒体出现之前，我们可能会认为，让人们更容易进行辩论会让我们团结起来，促进达成共识。而现实情况却恰恰相反：人们似乎比以往任何时候都更加愤怒，观点也更加两极分化。比如说，人们普遍认为社交媒体助长了2011年英国骚乱，以及最近在2020年美国总统大选后发生的冲击美国国会大厦事件。因此，人们迫切地想要了解是什么在推动这些分歧，以及可以采取什么措施(如果有的话)。事实

证明，物理学也许能给出一些答案。

将物理学应用于描述社会现象的想法至少可以追溯到17世纪哲学家Thomas Hobbes，他曾试图用伽利略的运动定律来描述社会的“物理现象”。近几十年来，建立与社会有关的物理模型已经对选举结果有了成功预测，揭示了两极分化观点如何变为主导，乃至给出消除分歧的方法，这些成就使其已经成为一个正式的研究领域，被称为“社会物理学”。

危险的相似

十多年来，政治和社会科学家将分裂的加剧归因于社交媒体泡沫和回声室效应(译者注：人们倾向接受与自己意见相似的观点，从而强化了偏见)，并认为缺乏与持不同观点的人交流会使我们的观点变得更极端。然而，当物理学家模拟社会行为，例如将相反观点视为相斥力时，却无法复现这种极化效应。事实上，这些模型甚至表明最终结果

会是达成更大的共识。

2020年，来自西班牙加泰罗尼亚理工大学和意大利CENTAI的物理学家Michele Starnini及其同事们试图通过为人们所持观点的强度与他们直接接触到的观点强度之间的关系建立函数模型，来找出推动两极分化的因素。早期的模型通常基于“建设性观点动力学”，即不受限制的交流最终会导致共识的达成，即使是在有争议的问题上。然而，他们的新模型引入了激进化动力学作为强化机制，这使得观点相似的人之间更有可能建立联系。Starnini的工作展示了极端观点是如何从温和的初始条件演变而来的。

研究模型得出了三种可能的状态：共识、极化观点或激进观点，最后一个状态指的是人们在所有话题范围内只接受某一端的观点。他们的分析结果符合我们的经验性观察，即当人们的观点受到他人的强烈影响，而所涉及的问题又存在争

议时，这些观点就会变得更加极端。再加上人们有与志同道合的人建立联系的强烈倾向，整个网络就不是激进化的，而是两极分化了(图1)。事实上，Starnini的团队发现，模型得出的观点极化分布符合对X(前身为推特)上参与堕胎、奥巴马医改和枪支管制辩论的用户的真实社交媒体数据分析，以及从脸书和油管等其他社交媒体网站获取的用户观点数据。

他们的结论看似显而易见。但是，在模型中复现已知行为是获得不那么直观的认识的第一步。比如说，Starnini及其同事们的分析表明，虽然观点的中间地带主要由活动程度较低的用户占据，但极端地带主要由最活跃的用户代表，因为他们的观点从志同道合的人那里得到了最有力的强化。

从极化到去极化——一种相变

社交媒体数据还为日益加剧的两极分化提供了其他见解。2023年1月，迈阿密大学统计物理学家宋超明(Chaoming Song)和他的学生Jiazhen Liu等人组成的研究小组比较了用户对于脸书上“硬”内容(政治)和“软”内容(体育和娱乐)相关帖子的观点强度。不出所料，他们发现硬内容的观点集中在两个极端，展现了一种两极分化的状态，而软内容的观点则更偏向于中间立场(图2)。

重要的是，宋和他的同事们发现，这些观点分布遵循一个标度定律。一共有三个稳定相：观点汇聚在中间立场的共识相；观点均匀分布在整個范围内，向极端位置略有增加，即“部分极化”；以及全面极化。在研究这些观点分布如何变化的关键问题时，将这些观点分布划

分为不同的相引入了一个有潜在价值的概念：相变。宋及其同事们证明了网络社会行为中的相变——就像凝聚态物理学中的相变一样——由序参量的临界点所表征，这里的序参量是最可能或最普遍的观点。

为了了解去极化转变的表现形式，Starnini及其同事们设计了另一个模型。该模型展示了相变的本质如何依赖于对某一个话题的观点是否与其他话题的观点相关。根据美国国家选举研究所的民意调查，研究人员发现了几种根据所表达的观点分布情况来看似乎相互关联的言论，比如“宗教为日常生活(提供)指导”和“如果同性伴侣违反了企业主的宗教信仰，企业主可以拒绝为其提供服务”。但是，同一项民意调查也提供了一些似乎不会引起相关反应的陈述，比如说“在美国出生的非法移民的子女应自动获得公民身份”和“美国应出兵打击伊

斯兰激进分子”。研究人员随后模拟了这些观点分布将如何随着网络中其他人的观点而演变，同时考虑到了“社会影响力”以及这些观点的顽固程度。

他们发现，当观点相关时，从极化相到去极化相(即共识相)的相变可能是平滑的，就像凝聚态材料中不同磁相之间发生的二级相变。另一方面，该模型表明，当观点不相关时，相变可能是突然的，就像从液体到气体的一级相变。

2015年，美国康奈尔大学的社会学家Daniel DellaPosta及其同事证明了政治党派越来越多地决定着个人在看似不相关的领域中的观点——从休闲活动和审美趣味，到食品消费和个人道德。举个简单的例子，你可以说自由主义者喜欢拿铁咖啡，而保守主义者喜欢猎鸟。如果我们的观点如DellaPosta及其同事所指出的那样越来越相关，可以

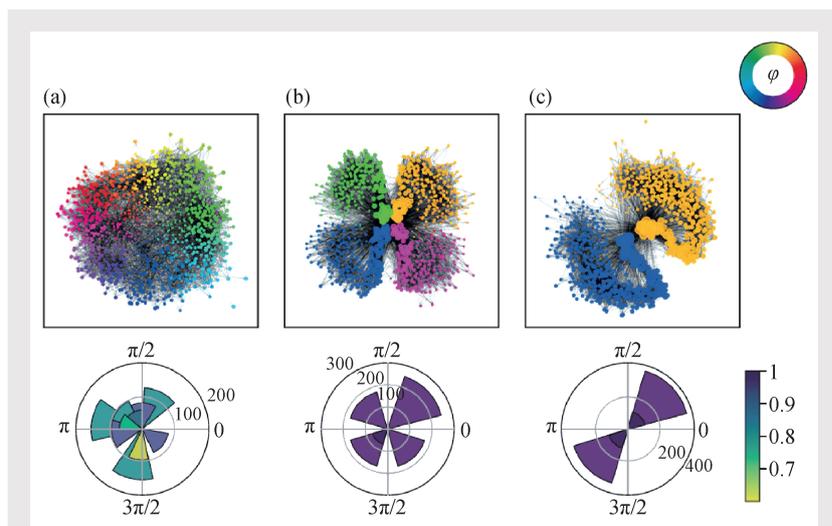


图1 社会结构。加泰罗尼亚理工大学的Michele Starnini及其同事们展示了社交媒体上的观点极化或分化现象。他们的模型指出了社交网络上三种不同动态状态下的观点：(a)接近共识状态；(b)无关联极化状态；(c)意识形态状态(在这种状态下，观点极化与表面上无关联的话题一致)。在网络的示意图(上图)中，每个节点根据其观点角度 ϕ 进行着色，其大小与信念强度成正比。社群在对应网络示意图下方的极柱图中表示，半径表示社群大小，颜色和宽度则对应于社群中所有个体对之间的平均余弦相似度。取向代表社群内所有对象的平均观点角度 ϕ 。占总节点数不到5%的小社群不予显示

期望从极化到非极化的转变更加平滑、更像二阶相变。

但究竟是什么首先引发去极化？而这又是否可取呢？有人可能会说，持有普遍一致观点的单一文化并不好，因为它不容易接纳任何多样性或对常规的偏离。根据 Della-Posta 的说法，社会政治理论家很早就发现争论有利于“在多元社会中存在的在某些问题上具有冲突和分歧的两派中能在其他问题上共事的交集群体”。那么，我们怎样才能建立一个更加宽容的社会呢？幸运的是，物理学的另一个领域提出了一些有趣的观点。

团结一致的干预措施

2023 年，华盛顿特区乔治·华盛顿大学的物理学家 Neil Johnson 及其同事提出了一个方程来模拟“反 X”网络社群的演变，其中 X 可以是任何东西，例如种族、宗教或民族。该方程与流体动力学中的一个基本方程，即描述波形演变的伯格

斯方程(Burgers' equation)，极为相似。这种联系很有道理，因为这些极端主义团体区别于温和团体的最常见特征就是他们是作为“流氓波”出现的。无论极端分子的观点如何，正如 Johnson 所说，他们似乎都是“突然冒出来的”。这很可能是因为他们长期在“雷达”探测不到的地方活动，然后突然在主流社会中引起关注。

在 Johnson 及其同事们的研究中，他们将个人或群体看做一个个矢量，这样不同的个体经历可以在矢量的多个维度上体现出来。个人加入一个群体或群体融合形成一个更大群体的概率，就取决于他们对应的矢量方向有多么一致(图 3)，研究人员称之为“互联网上集体互相吸引”。他们的模型表明，减缓反 X 群体突然出现的诀窍是让更多不同的用户涌入网络，这样同类矢量就不那么容易找到对方。这个建议与使用算法阻止另类用户参与的做法相悖，因为这些

算法实际上会降低多样性。

同样，过于主动地让社交媒体用户接触反对观点可能适得其反。2018 年，由美国杜克大学社会学家 Chris Bail 及其同事们领导的一项研究表明，这种刺破回声室泡沫的努力可能会加剧两极分化。次年，Johnson、宋等人指出，脸书的某些算法更新可能会产生同

样的反作用。尽管这些更新旨在建立用户之间的新联系并培养新社群，但他们认为，这些算法更新可能导致孤立的极端群体出现，最终会造成社群网络断断续续地演化并且很容易决裂和破碎。

然而，在现实世界中与持不同观点的人进行对话是有益的。我们每天邂逅的人不可能事事都与我们持相同的观点，但是尽管在某个问题上可能有观点冲突，我们总能在其他问题上达成一致观点。2022 年，荷兰阿姆斯特丹大学的计算社会科学家 Petter Törnberg 的一项研究表明，在虚拟世界中，社交媒体无法自然产生这种多样却稳定的差异“拼凑”，而是会扩大我们的潜在接触人群，使得志同道合的人之间产生联系变得更容易。由于没有内部的调节机制，人们的观点更难融合，更具党派性。

英国伦敦国王学院信息学博士生 Luke Thorburn 和前计算机科学家、现致力于研究新兴技术对社会影响的 Aviv Ovadya，正在尝试开发能产生较少分歧的社交媒体平台。他们正在开展一个名为“桥接系统”的项目，旨在寻找超越传统“点赞”和“转发”的方式来对社交媒体帖子进行排名的方法。这个项目的想法不是盲目地让人们直接面对完全相反观点的帖子，而是推荐带有一部分共同观点糖衣的不同观点的帖子(译者注：即加入表面活性剂以降低界面能)，即 Thorburn 所说的“多元认同”。例如，你可能不会轻率地邀请一个喝拿铁咖啡的自由派与一个猎鸟的保守派建立联系，但是你可以基于超越常规党派界限的共同兴趣，比如凝聚态物理学或某种体育运动，邀请两个政治立场相反的人进行交流。

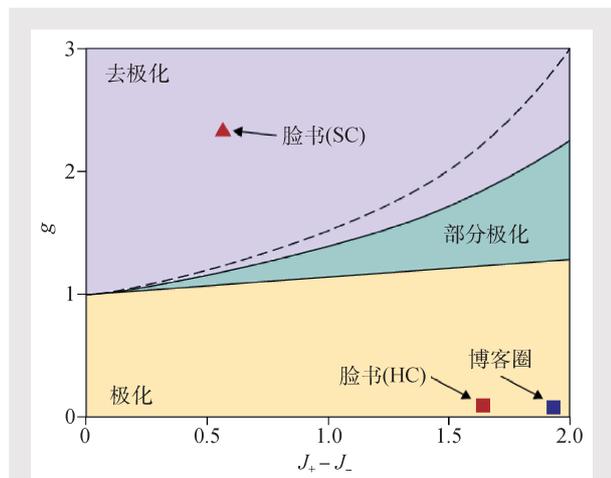


图2 相图。迈阿密大学的宋超明及其同事展示了如何用相变来描述观点分布的演变。在此相图中，参数 g 是意见一致的速率乘以创建链接与破坏链接的比率。参数 J_+ 是志同道合者倾向于建立链接(结党)的强度， J_- 是观点相左者倾向于断开链接(绝交)的强度。红色正方形表示脸书上的硬内容(HC)，例如政治；红色三角形代表软内容(SC)，例如体育和娱乐

社交媒体之外

像物理学中的任何领域一样，检验一个好模型的关键不仅在于解释现有数据，还要能做出正确的预测，而这正是物理学家 Serge Galam 的工作如此引人注目的原因。现就职于巴黎政治学院的 Galam 从 20 世纪 70 年代社会物理学的出现以来就一直致力于对其的研究，当时这一领域遭到了物理学家强烈的反对。他说：“原子从未让我感到兴奋，但人类会！”

他的模型采用“少数服从多数”的规则来更新群体的观点，使他们的意见逐渐倾向于多数派，就像每天在咖啡时间进行的小组聊天可能会逐渐改变一个没有内在偏见的人一样。毕竟，保持与周围人一致的立场更让人省心。Galam 认为这借鉴了物理领域中到处都有体现的能量最小化原则，比如磁体中的自旋排列或生物细胞中的蛋白质折叠。

Galam 考虑的是离散观点，即重点在于观点的方向，而非观点的强度，比如在选举中的投票。在群体观点的演变中，他将每一方观点的最终比例称为“吸引子”。对于一个趋向共识的简单系统，存在两个吸引子：所有人同意某一观点，或者所有人都同意其相反观点。结果取决于观点的初始比例，而观点初始的 50:50 混合是临界点，在这个临界点上最终结果可能会朝任何一方倾斜。要发生极化，一个已经达成共识的群体必须与另一个达成相反共识的群体接触。

但是，达成共识的道路并不总是一帆风顺的——在两种选择似乎同样可接受的情况下人们会游移不定。然后，人们通过无需论证的

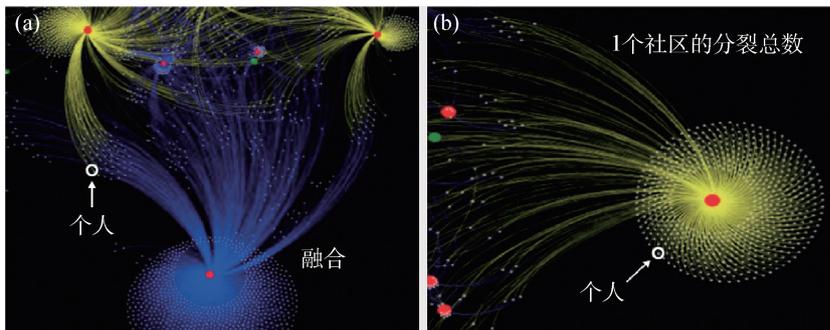


图3 社群聚类和集体化学。乔治华盛顿大学的 Neil Johnson 及其同事们在俄罗斯社交网站 VKontakte 上观察到第 t 天(黄色)和第 $t+1$ 天(蓝色)之间反美仇恨社群的融合(a)和分裂(b)。红色节点代表后来被关闭的反美社区(完全裂变)，而绿色节点代表截至 2023 年 6 月仍未关闭的反美社群。黄色链接指向第 $t+1$ 天从反美社群中移除的个人(白点)，蓝色链接指向在第 $t+1$ 天加入到反美社群的个人。图(a)和(b)的空间分布是使用 ForceAtlas2 映射的更完整网络的特写，这意味着出现在更近处的节点相互联系更紧密。图(b)还表明，在完全裂变时，该社群中很少有个体同时也是其他社群的成员

“偶然”选择进行决策。在这里，Galam 假设“偶然”选择实际上受到群体中最强偏见的影响，这反过来又复现了他所称的“民主少数派传播”，即一个群体中最强的偏见，而不是人数最多的观点，左右最终结果。事实上，Galam 认为这种偏见是 2016 年美国总统选举的决定性因素。除了偏见，Galam 的模型还包括其他心理因素，比如逆反心理和固执己见，这两种因素都能改变吸引子和临界点。

Galam 指出，如果在理论模型中加入的心理因素是固执，模型将演化成一个僵硬的观点极化系统，各个群体间具有仇恨与排斥的氛围。相反，如果加入更多的逆反心理，支持两方极化观点的人数比例将会改变，气氛会变得更加宽容和接纳，即使最初双方的人数比例与加入固执时的相同。在 2023 年的一项分析中，Galam 从熵的角度描述了上面的液化极化与凝固极化之间的区别，强调了群体中个体的性质比选民的分布更重要。

事实上，只要有足够多的逆反者，就不会出现临界点：结果总是趋于两种观点的比例为 50:50，没有达成共识的可能。如果固执者足够多，临界点也会消失，但这次吸引子是向多数派的观点倾斜，而不是 50:50 的比例。Galam 说：“这导致了关于竞选获胜策略的一种令人不安的，甚至不道德的观点的出现。”要赢得选举，候选人无需努力说服最多的选民，而只需说服少数最固执的选民，并使他们固执地支持自己。

无论是用选票还是钱包投票，这些认识都很适用，并且可能在市场营销、广告和类似的操纵活动上具有强大的威力，因为了解你的竞选目标——净多数的关键影响者——是成功的一半。然而，Galam 强调，它也可以用于防止“僵局”，特别是考虑到当今社会动态的困境，比如假新闻。

“这就像任何科学发现一样，”他说，“它可以被用于好的方面，也可以被用于坏的方面。”