arXiv:2302.01919v1 [math.PR] 3 Feb 2023

集体行为中的临界数目

Jacob Calvert

摘要。自然集体中的个体可能并不知道自身的数量，但这些集体仍可能表现出对数量高度敏感的行为。本文证明，数量无感知个体的集体行为甚至可以具有一个临界基数，在该临界值之上和之下，集体行为会呈现质的差异。我们以一族零一律的形式对临界基数这一概念进行形式化，并引入一种称为链式激活与输运（CAT）的集体运动模型，该模型具有临界基数。

CAT 将 $n \geq 2$ 个体的集体运动描述为一个马尔可夫链，该链每次重新排列 d维网格中由 n个元素组成的子集，一次处理 $m < n$ 个元素。根据个体的动力学，在每一步中，CAT 都会从集合中移除 m 个元素，然后逐步将 m 个元素添加到剩余部分的边界上，其方式倾向于连续添加和移除彼此邻近的元素。本文证明，如果 $d \geq 3$，那么就其直径的行为而言，CAT 具有 $n_c = 2m + 2$ 的临界基数。具体而言，如果 $n < n_c$，那么这些元素会形成一个“簇”，其直径具有几乎处处有限的下极限。然而，如果 $n \geq n_c$，那么在一个几乎处处有限的时刻，集合将由若干个包含 $m+1$ 至 $2m+1$ 个元素的簇组成；此后这些簇将不断彼此远离，从而导致直径无限增长。

临界个体数的存在意味着，集体能够表现出仅由其数量决定的“相变”，而非由密度或相互作用强度等因素决定。这一事实挑战了关于集体行为的普遍认知；引发了有关数量在自然集体行为中所起作用的基础科学问题；并为可编程物质（如机器人群、智能材料和合成生物系统）带来了新的功能设想。更广泛地说，这表明我们可以通过研究重排有限集合的随机过程，探索自然集体和人工集体可能呈现的行为。

目录

1. 引言 2
2. 预备知识 3
3. 证明思路 4
4. 直径坍缩 5
5. 直径增长 6
6. 簇形成 7
7. 不可约性 参考文献 17
18
27
31
33

电子邮箱地址：jacob_calvert@berkeley.edu。2020数学主题分类。60J10、60K35 和 82C22。关键词和短语。马尔可夫链、临界基数、集体行为、可编程物质。

1. 引言

任何集体行为都面临一个根本问题：集体的数量必须达到多少，才能表现出这种行为？特别是，是否存在一个临界基数[1]，集体数量高于该值时会表现出这种行为，而低于该值时则不会？实验依据表明，自然集体在某些重要行为方面可能具有临界个体数（表 1）。这一可能性非同寻常，因为这些集体的组成个体未必知道自身的数量，尽管它们或许能够推断出这一数量[Pra05]。为探讨这些问题，本文将临界个体数的概念形式化，并提出一种集体运动模型，该模型可证明具有临界个体数。

表 1. 六项表明临界个体数存在的实验（n_c）。行为的精确定义及 n_c 的建议值取决于实验细节。

集体类型	行为	建议的 n_c	参考文献
法老蚁群落 (法老蚁)	觅食是基于路径的	≈ 700	[BSR01]
摇蚊群 （摇蚊）	每只摇蚊的体积已达到饱和	≈ 10	[PO14]
切叶蚁群落 (*Acromyrmex versicolor*)	种群随之增长真菌培养物质量	≈ 89	[CF14]
水分子簇	形成冰I	≈ 90	[MBD+19]
慈鲷鱼群 (*Etroplus suratensis*)	最有可能的状态是各向同性运动	≈ 60	[JMAK+20]
火蚁群 (*Solenopsis invicta*)	形成稳定的筏状群体	≈ 10	[KHKH22]

临界个体数的存在意味着，集体可以表现出由纯粹由个体数量而非例如其密度或相互作用的性质或强度所介导的行为相变——也就是突然而显著的变化。这一事实挑战了关于集体行为的传统观点。例如，Ouellette 和 Gordon 写道：“ [一个集体] 的属性和功能源于个体之间的相互作用。这意味着，没有哪个个体对于群体的运作是必不可少的。” [OG21]然而，如果集体的某种行为处于临界个体数，那么每一个个体都是必不可少的。出于同样的原因，临界个体数的存在表明，可编程物质（例如机器人群）可能具有新的功能。事实上，一种经过实验验证的物质编程方法采用了受伊辛模型启发的相变：通过增强单个机器人之间的吸引相互作用，机器人群会从分散相转变为能够实现集体运输的紧凑相 [LDC+21]。类似地，临界个体数的存在可能使可编程物质能够根据其数量表现出不同的功能。

1.1. 记号。 关于 $\mathbb{Z}$ 和 $\mathbb{R}$。对于 $i,\ j \in \mathbb{Z}$，我们记 $\mathbb{Z}_{\geq i} = \{i, i+1, \dots\}$ 和整数区间 $[\![i,\ j]\!] = \{k \in \mathbb{Z} : i \leq k \leq j\}$。特别地，我们记为 $\mathbb{N} = \mathbb{Z}_{\geq 0}$。我们用 $\mathbb{R}_{\geq 0}$（$\mathbb{R}_{>0}$）表示非负（正）实数。我们用 $[r]$ 表示 $r \in \mathbb{R}$ 的整数部分。

关于 $\mathbb{Z}^d$。对于 $d \in \mathbb{Z}_{\geq 1}$ 和 $A,\ B \subset \mathbb{Z}^d$，我们定义 $\operatorname{diam}(A) := \sup_{x,\ y \in A} \|x - y\|$，以及 $\operatorname{dist}(A, B) := \inf_{x \in A,\ y \in B} \|x - y\|$，其中 $\|\cdot\|$ 是欧几里得范数。对于 $A \subset \mathbb{Z}^d$，我们将边界和闭包记为

[1]我们使用“数量”一词，因为它比“数”更具描述性，而且与“大小”不同，它不会与集体的空间范围相混淆。

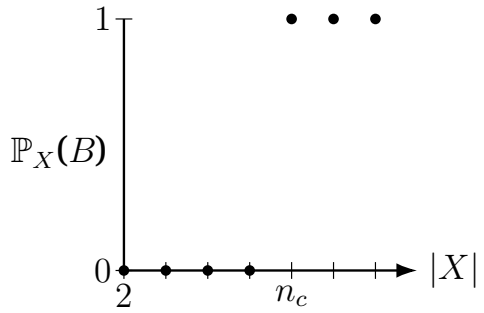

图1. 状态空间 $\mathcal{S}$ 上的随机过程 $\mathscr{X} = (\mathscr{X}_0, \mathscr{X}_1, \dots)$ 满足（1），若对于每个 $X \in \mathcal{S}$，该图均成立，则其相对于事件 $B \in \sigma(\mathscr{X})$ 的临界基数为 n_c。

A 的 ∂A 为 $= \{x \in \mathbb{Z}^d \setminus A : \operatorname{dist}(x, A) = 1\}$ 和 $\bar{A} = A \cup \partial A$。我们将以 $x \in \mathbb{Z}^d$ 为中心、半径为 $r \in \mathbb{R}_{>0}$ 的离散欧几里得球记为 $\mathcal{B}_x(r) := \{y \in \mathbb{Z}^d : \|x - y\| < r\}$。对于 $i \le j$，我们用 $L_{i,j}$ 表示线段 $\{ie_1, \dots, je_1\}$，其中 e_j 表示 $\mathbb{R}^d$ 的 j^{th} 标准单位向量；我们将这类线段的平移像的集合记为 $\mathcal{L} := \{L_{1,k} + x : k \in \mathbb{Z}_{\ge 1}, x \in \mathbb{Z}^d\}$。我们将与 $x \in \mathbb{Z}^d$ 在 $A \subseteq \mathbb{Z}^d$ 中连通的元素集合定义为 $\operatorname{comp}_A(x)$。特别地，只要 $x \notin A$，就有 $\operatorname{comp}_A(x) = \emptyset$。

其他。对于 $k \in \mathbb{Z}_{\ge 1}$，我们用 $A \subset_k B$ 表示 A 是 B 的一个 k 元素子集。此外，我们用 $f \lesssim g$ 表示“通用”的正数 c 的估计值 $f \le cg$，并用 $O(g)$ 表示这样的量 f。对于反向估计，我们类似地使用 $f \gtrsim g$ 和 $\Omega(g)$。最后，对于状态空间 $\mathcal{S}$ 上的随机过程 $\mathscr{X} = (\mathscr{X}_t)_{t \in \mathbb{N}}$，我们用 $\mathbb{P}_X$ ($\mathbb{E}_X$) 表示在给定 $X \in \mathcal{S}$ 满足 $\mathscr{X}_0 = X$ 的条件下，$\mathscr{X}$ 的条件分布（期望）；并用 $\sigma(\mathscr{X})$ 表示 $\mathscr{X}$ 生成的 σ 域。

1.2. **临界基数的定义**。我们将个体的状态建模为 $\mathbb{Z}^d$ 的一个元素，将集体的状态建模为 $\mathcal{S} \subseteq \{S \subset \mathbb{Z}^d$ 中的一个元素：$2 \le |S| < \infty\}$。集体的动力学是定义在 $\mathcal{S}$ 上的随机过程 $\mathscr{X} = (\mathscr{X}_0, \mathscr{X}_1, \dots)$，它保持个体数量不变：

$$\forall X \in \mathcal{S}, \quad \mathbb{P}_X(|\mathscr{X}_0| = |\mathscr{X}_1| = \cdots) = 1. \tag{1}$$

我们根据一族零一律来定义临界基数（图1）。

定义1.1（临界基数）。我们称 $\mathscr{X}$ 相对于 $B \in \sigma(\mathscr{X})$ 具有 $n_c \in \mathbb{Z}_{\ge 2}$ 的临界基数，若

$$\forall X \in \mathcal{S}, \quad \mathbb{P}_X(B) = \begin{cases} 0 & |X| < n_c, \\ 1 & |X| \ge n_c. \end{cases} \tag{2}$$

定义1.1旨在将这样一种行为概念形式化：当且仅当由 $\mathscr{X}$ 建模的集体具有 n_c 个或更多组成个体时，它才表现出行为 B。虽然这一界定足以满足我们的目的，但我们注意到，其范围可以扩展到 $\mathbb{Z}^d$ 子集上的离散时间随机过程之外；此外，也可以用其他方式将“行为”和“表现出”形式化。例如，由于行为 B 可以是 $\sigma(\mathscr{X})$ 中的任意事件，定义1.1允许出现像 $\{|\mathscr{X}_0| \ge n_c\}$ 这样刻意构造的行为，它通过明确提及数量而满足条件（2）。[2] 另一种做法是不允许 $\sigma(\cup_{t \in \mathbb{N}} |\mathscr{X}_t|)$ 中的行为。至于“表现出”，还有一种替代方案

[2] 同样，由于对 $\mathscr{X}$ 的动力学没有任何限制，一种人为设计的动力学（例如仅当 $|\mathscr{X}_0| \ge n_c$ 时才是非平凡的动力学）即使在所讨论的行为完全不涉及数量的情况下，也能满足（2）。

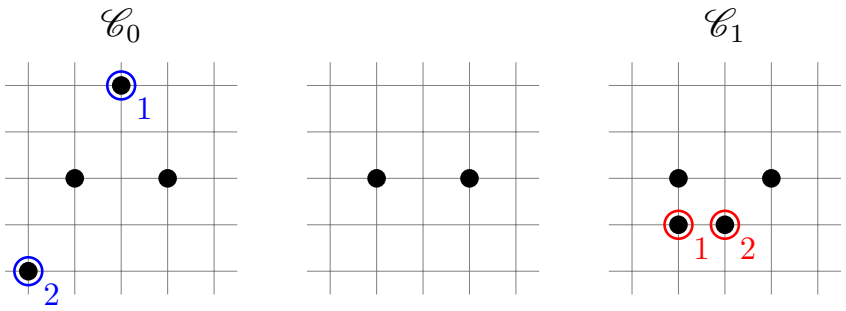

图 2. 参数为 $(d,\ m,\ n) = (2,\ 2,\ 4)$ 时动力学的一个示例。激活发生在以蓝色圆圈标出的元素处（左图），由此得到一个中间集合（中图）。随后，传输至以红色圆圈标出的元素处 $\mathscr{C}_1 = \mathscr{C}_{0,\ 4}$（右图）。

是将条件 $\mathbb{P}_X(B) = 1$ 当 $|X| \geq n_c$ 时成立，替换为较弱的条件：$\mathbb{P}_X(B) > 0$ 当 $|X| \geq n_c$ 时成立。[3]

1.3. 链式激活与输运。 本文的主要结果涉及一种称为链式激活与输运（CAT）的集体运动模型，它属于对图的有限子集进行重新排列的马尔可夫链这一类，例如调和激活与输运 [CGH21、Cal21], 竞争侵蚀[GLPP17、GP18、GLS19], 以及可编程物质的几何阿米巴机器人模型 [DDG+14、CDRR16、AACD+17、SCD+18、CDG+19、LDC+21]。具体而言，CAT 将定义在状态空间 $\{S \subset \mathbb{Z}^d : 2 \leq |S| < \infty\}$ 上的马尔可夫链加以推广；该链在每一步中，先从集合中移除一个均匀随机选取的元素，再向剩余集合的边界添加一个均匀随机选取的元素。CAT 的基础数据如下：

- $d \geq 1$，底层网格的维数；
- $m \geq 1$，每步移动的元素数量；以及
- $\beta \in \mathbb{R}$，控制元素倾向于移动多远的参数。

CAT 是定义在状态空间 $\mathcal{S} = \{S \subset \mathbb{Z}^d$ 上的马尔可夫链 $(\mathscr{C}_t)_{t\geq 0} : m + 1 \leq |S| < \infty\}$ 每步从集合中移除（激活）m 个元素，然后将（传输）m 个元素逐个添加到集合的边界上，添加概率取决于连续元素之间的距离。换言之，共有 $2m$ 个子步骤；设 $\mathscr{C}_{t,l}$ 表示时刻 t 经过 l 个子步骤后得到的集合，则有 $\mathscr{C}_{t+1} = \mathscr{C}_{t,2m}$（图 2）。

我们在 $\mathscr{C}_t$ 中均匀随机地激活第一个元素。然后，给定 x_l 是被激活或传输的第 l^{th} 个元素，$(l+1)^{\text{st}}$ 子步骤按以下两种方式之一进行。如果 $l < m$，那么我们激活

$$x_{l+1} \in \mathscr{C}_{t,l} \text{ with probability } \propto \|x_l - x_{l+1}\|^{-\beta}.$$

否则，若 $l \geq m$，则我们传输到

$$x_{l+1} \in \partial\mathscr{C}_{t,l} \text{ with probability } \propto \|x_l - x_{l+1}\|^{-\beta}.$$

事实上，为了处理当 $l = m$ 时 $x_l = x_{l+1}$ 的可能性，我们用 $\phi(x_l, x_{l+1})$ 代替 $\|x_l - x_{l+1}\|^{-\beta}$，其中

$$\phi(u, v) := \max\{\|u - v\|, 1\}^{-\beta} \tag{3}$$

对于 $u,\ v \in \mathbb{Z}^d$。第 2.1 节给出了 CAT 转移概率的一个公式。

[3]虽然这些条件通常并不等价，但若 $\mathscr{X}$ 是 $\mathcal{S}$ 上的常返马尔可夫链，且 B 属于 $\mathscr{X}$ 的尾 σ-域，那么 Blackwell–Freedman 零一律 [BF64，定理 1]表明它们是等价的。

我们强调，尽管在 CAT 中激活以均匀随机的方式开始，但这并不意味着 CAT 要描述集体运动，个体就必须知道其数量。事实上，我们可以将 CAT 视为描述数量无感知个体的集体运动：这些个体根据独立同分布指数时钟响起的时刻异步迈步 [LDC+21]。关于动力学中的非局部组成部分，我们可以将 CAT 视为描述对自适应长程相互作用作出响应的个体，例如声学相互作用 [GIP+16、GG17、GPvdV+20]、视觉相互作用 [PMRT14、LWBB19、BR20、HAC+21]，或趋化性相互作用 [DVM04] 。

从 CAT 的定义可以看出其四个重要性质。

1. 数量守恒。CAT 在（1）的意义下保持元素数量不变。 .

2. 任意大的直径。CAT 构型可以包含多个连通分量，因此元素数量不会限制构型的直径。具体而言，如果构型C 至少包含三个元素，那么对于任意 $r>0$， $\mathscr{C}_t$ 的直径最终会超过 r, $\mathbb{P}_C$–几乎必然。

3. 直径的非对称行为。由于元素必须被传输到其他元素的边界，构型的直径每一步至多增加 m 。换言之，CAT 具有至多线性增长（a.m.l.g.）性质

$$\mathbb{P}_C(\operatorname{diam}(\mathscr{C}_{t+1}) \le \operatorname{diam}(\mathscr{C}_t) + m \text{ for all } t \ge 0) = 1, \tag{4}$$

对于每个构型 C。与之相反，直径可能会骤然减小。例如，如果 $n = m+1$且 $C \subset_n \mathbb{Z}^d$ 的直径为 $r \ge n$，那么第一步中直径至少会减小 $r-n$， $\mathbb{P}_C$–几乎必然。

4. 平移不变性。由于其转移概率仅取决于元素之间的距离，CAT 满足

$$\mathbb{P}_C(\mathscr{C}_1 = D) = \mathbb{P}_{C+x}(\mathscr{C}_1 = D + x)$$

对于任意 $C, D \in \mathcal{S}$ 和 $x \in \mathbb{Z}^d$。对于每个构型 C ，我们将由其平移像构成的等价类记为:

$$\widehat{C} = \left\{D \subset \mathbb{Z}^d : C = D + x \text{ for some } x \in \mathbb{Z}^d\right\}. \tag{5}$$

1.4. 主要结果陈述。 下文中，我们假设 $d \ge 3$ 和 $\beta > 2$。我们的主要结果是，CAT 关于其直径的行为存在一个临界基数（图3）。通过选择 m，可以使这个临界基数任意大。

定理 1.1 （临界基数）。 对于事件$\{$直径$(\mathscr{C}_t) \to \infty\}$，*CAT* 的临界基数为 $n_c = 2m+2$。

对于 CAT 在 n_c之上和之下所呈现行为的定性差异，我们还能证明更多内容。这是接下来三个定理的主题。第一个结果指出，当 $n < n_c$时，CAT 的直径会发生坍缩；也就是说，无论初始集合的直径多大，其直径通常会在 n 步内降至 n 的某个函数以下。

定理 1.2 (直径坍缩**)。** 若 $n < 2m+2$，则存在 $p > 0$ 使得

$$\mathbb{P}_C\left(\operatorname{diam}(\mathscr{C}_n) \le 2n^2\right) \ge p, \tag{6}$$

对每个 $C \subset_n \mathbb{Z}^d$。

这一结果表明，若 $n < n_c$，则 $\mathscr{C}_t$ 的直径几乎必然会无限多次不超过 n 的某个函数。因此，在这种情况下，直径的下极限几乎必然是有限的。定理 1.2 的关键特点是，式 (6) 中的步数、直径上界和概率下界都不依赖于直径

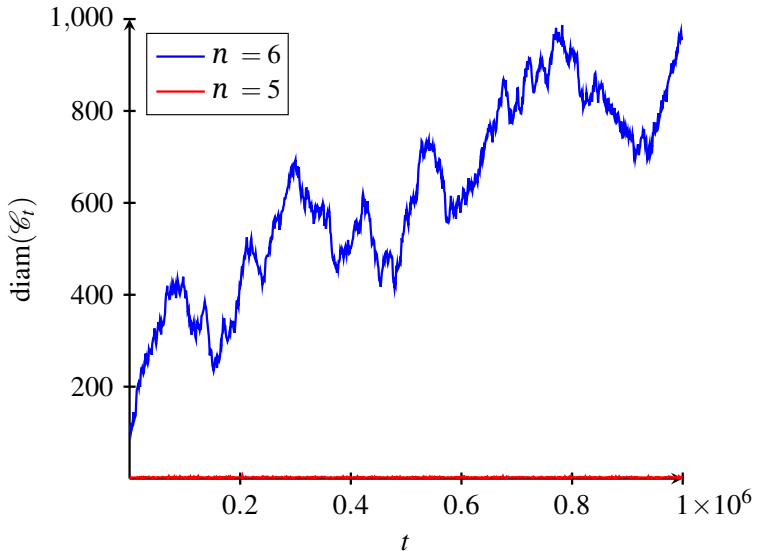

图3. CAT 的代表性模拟，参数为（d，m，β）=（3，2，4），经过 10^6 步，从直径为 100、包含 n 个元素的集合开始。

C；除此之外，它们的具体取值并不重要。事实上，如果 p 可以依赖于 diam(C)，那么只需对转移概率作简单估计，就足以证明式 (6)，而无须假设 n_c。

如果 $n \geq n_c$，那么 $\mathscr{C}_t$ 最终会达到一种构型，其中包含彼此充分分离的“簇”，这些簇具有足够数量的元素，因而能够以某种意义持续存在。具体而言，这些簇包含足够多的元素（至少$m+1$），从而避免在单一步数内被完全激活；但元素又不够多（至多 $2m+1$），无法分裂成能够持续存在的更小簇。

为使其精确定义，我们将 CanPersist$_r$ 对于 $r \geq 0$ 定义为 $\mathcal{S}$ 的子集，其中包含那些存在一个划分 $(V^j)_{j=1}^J$ 的集合，使得

$$\forall j \in [\![1, J]\!], \quad V^j \in \mathcal{L}, \quad |V^j| \in [\![m+1, 2m+1]\!] \quad \text{and} \quad \operatorname{dist}(V^j, V^{\neq j}) \geq r, \tag{7}$$

其中 $V^{\neq j} = \cup_{i \neq j} V^i$。为便于使用以下结果，我们规定每个 V^j 都属于 $\mathcal{L}$，即与e_1平行的线段集合。

定理 1.3（簇形成）。 存在 $p > 0$ 使得

$$\mathbb{P}_C(\mathscr{C}_{2n^2 r} \in \mathsf{CanPersist}_r) \geq p^r, \tag{8}$$

对于每个 $C \subset_n \mathbb{Z}^d$ 和 $r \in \mathbb{Z}_{\geq 2n}$。

定理 1.3 指出，CAT 需要约 r 步，才能到达一种状态：其中只包含彼此分离 r 的簇，并且这些簇能够持续存在；这一结果以至少依赖于 r 的概率成立。与定理 1.2 一样，定理 1.3 的关键特点是，(8) 中没有出现 C 的直径。如果这些簇彼此充分分离，那么由于它们具有持续存在所需的适当元素数量，它们往往会像 d 维随机游走一样移动，因此经过 t 步，其分离程度往往会增长至 $t^{1/2}$。随着它们变得越来越分离，簇之间交换元素的可能性也越来越低，从而促进分离程度进一步增长，如此循环。这一机制是 $n \geq n_c$ 时直径增长的基础。

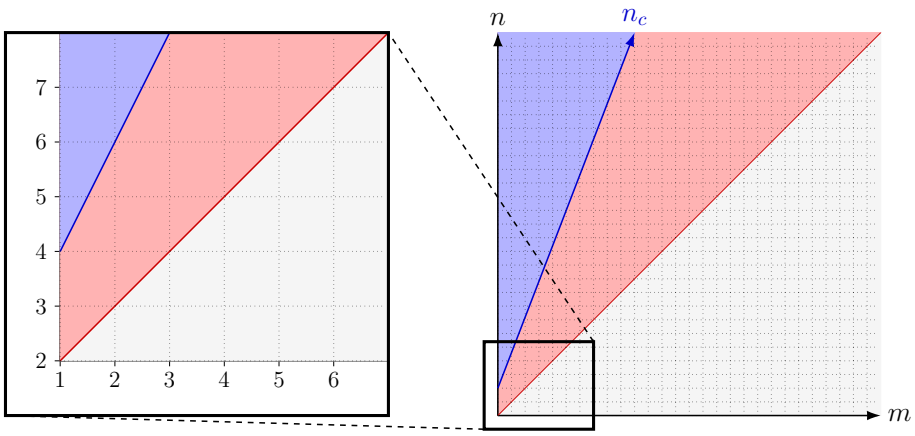

图 4。CAT 在（m, n）网格中的相图包含直径持续增长（蓝色）和频繁直径坍缩（红色）的区域，两者由临界个体数（标记为 n_c 的线）分隔。CAT 在 $n \leq m$ （灰色）时未定义。

定理 1.4 (直径增长)。 如果 $n \geq 2m+2$，那么存在 $\gamma > 0$，使得对于任意 $\delta \in (0, 1/2)$，都存在 $p > 0$ 和 $s \in \mathbb{Z}_{\geq 0}$，使得

$$\mathbb{P}_C\left(\operatorname{diam}(\mathscr{C}_t) \geq \gamma(t-s)^{\frac{1}{2}-\delta} \text{ for every } t \geq s\right) \geq p, \tag{9}$$

对每个 $C \subset_n \mathbb{Z}^d$。

图 4 总结了定理 1.2 和定理 1.4，它们推出了定理 1.1。

定理 *1.1* 的证明。 记 $B = \{\operatorname{diam}(\mathscr{C}_t) \to \infty\}$ 和 $n_c = 2m+2$。定理 1.2 和博雷尔–坎泰利引理推出，若 $n < n_c$，则 $\operatorname{diam}(\mathscr{C}_t)$ 的下极限有限 $\mathbb{P}_C$–几乎必然成立，因此对每个 $C \subset_n \mathbb{Z}^d$，$\mathbb{P}_C(B) = 0$。然而，若 $n \geq n_c$，则定理 1.4 和博雷尔–坎泰利引理推出，对每个 $C \subset_n \mathbb{Z}^d$，$\mathbb{P}_C(B) = 1$。□

我们的主要结果中的最后一项表明，CAT 的临界数量恰好对应于从$\widehat{\ }$ $\mathscr{C}_t$ 的常返性到其暂态性的相变。注意，要使 CAT 形成一个集合 U，其中 $\mathscr{C}_t$ 对于某个 $t \geq 1$成立，必须使 U 具有一个渐进边界，即包含一个满足下述条件的 m-元素子集 $\{x_1, \ldots, x_m\}$：

$$x_j \in \partial\big(U \setminus \{x_j, \ldots, x_m\}\big) \quad \text{for } j \in [\![1, m]\!]. \tag{10}$$

特别地，如果 $m = 1$，那么对于任何不完全由孤立元素组成的集合 U，式 (10) 都成立（即，元素在 U 中没有邻居）。我们将对 m 使用归纳法证明，这一条件也是充分的，即 CAT 在此类集合的集合上是不可约的。

定义 1.2 （渐进边界）。 如果 U 中存在一个满足（10）的 m-元素子集 $\{x_1, \ldots, x_m\}$，则集合 $U \in \mathcal{S}$ 具有渐进边界。我们用 $\operatorname{Prog}_{m,\, n}$ 表示具有渐进边界的 n-元素子集 $\mathcal{S}$ 的集合。我们用 $\operatorname{Prog}_{m,\, n}$ （5） 表示这类集合的等价类集合，其中每个$\widehat{\ }$等价类由具有渐进边界的 n-元素集合的所有平移像组成。

$\widehat{\ }$**定理 1.5** （从正常返性到暂态性的相变）。若 $n < 2m+2$，则 $\mathscr{C}_t$ 存在唯一的平稳分布，其支撑集为 $\widehat{\ }\operatorname{Prog}_{m,\, n}$，并且从 $\mathbb{Z}^d$ 的任意 n 元子集出发，都会收敛到该分布。然而，若 $n \geq 2m+2$，则它是暂态的。

标准马尔可夫链理论从不可约性、非周期性和正常返性三个方面刻画唯一平稳分布的存在性。我们在不同的命题中分别证明这些性质。

命题 1.1 (不可约性)。 $\mathscr{C}_t$ 在$\mathrm{Prog}_{m,n}$ 上不可约，对每个 $n > m$均如此。

我们将在第 7 节证明不可约性。关键步骤表明， $\mathscr{C}_t$ 可以通过按元素数量进行归纳，从一条直线上构造出任意构型。 $\mathscr{C}_t$ 具有非周期性这一事实证明起来更简单；我们在此陈述并证明它。

命题 1.2(非周期性)。 $\mathscr{C}_t$ 在$\mathrm{Prog}_{m,n}$ 上是非周期的，对每个 $n > m$均如此。

证明。 假设 $C = \{je_1 : j \in [\![1, n]\!]\}$。注意， $C \in \mathsf{Prog}_{m,n}$ ，因为令$y_j = je_1$，即可满足（10），其中 $j \in [\![1, m]\!]$。它是非周期的，因为 $\mathbb{P}_C\big(\widehat{\mathscr{C}_1} = \widehat{\mathscr{C}_0}\big) \geq \mathbb{P}_C\big(\mathcal{A}_0 = \mathcal{T}_0 = y\big) > 0$. 非周期性是一种类性质，因此命题 1.1 蕴含 CAT 在 $\mathsf{Prog}_{m,n}$上是非周期的。□

命题 1.3 (正常返性)。 若 $n < 2m + 2$，则 $\mathscr{C}_t$ 在 $\mathrm{Prog}_{m,\ n}$ 上正常返。

命题 1.3 可由定理 1.2 推出，因为定理 1.2 表明 $\mathscr{C}_t$ 的直径是一个李雅普诺夫函数。

命题 1.4 (福斯特–李雅普诺夫定理；FMM95] 中的定理 *2.2.4*)。 一个不可约马尔可夫链$(X_t)_{t\geq 0}$ 在可数状态空间 $\mathcal{V}$ 上正常返，当且仅当存在

$$f : \mathcal{V} \to \mathbb{R}_{>0}, \quad g : \mathcal{V} \to \mathbb{Z}_{\geq 1}, \quad \epsilon > 0, \quad \textit{and} \quad \textit{finite} \ \ \mathcal{U} \subset \mathcal{V}$$

使得

$$\forall U \in \mathcal{U}, \quad \mathbb{E}_U\big[f\big(X_{g(U)}\big)\big] < \infty, \quad \textit{and} \tag{11}$$

$$\forall U \notin \mathcal{U}, \quad \mathbb{E}_U\big[f\big(X_{g(U)}\big)\big] \leq f(U) - \epsilon g(U). \tag{12}$$

命题 *1.3* 的证明。 考虑由 $X_t = \mathscr{C}_t$ 在状态空间 $\mathcal{V} = \mathsf{Prog}_{m,\ n}$上定义的马尔可夫链，其中 $n < 2m + 2$，并采用如下选择

$$f(V) = \mathrm{diam}(V), \quad g(V) \equiv n, \quad \epsilon = 1, \quad \text{and} \quad \mathcal{U} = \{V \in \mathcal{V} : f(V) \leq b\},$$

对一个尚未指定的 $b > 0$。根据命题 1.1，$(X_t)_{t\geq 0}$ 在 $\mathcal{V}$ 上不可约，而该空间是可数的。因此，根据命题 1.4， $\mathscr{C}_t$ 在 $\mathsf{Prog}_{m,\ n}$ 上正常返，当

$$\forall U : \mathrm{diam}(U) \leq b, \quad \mathbb{E}_U\big[\mathrm{diam}(\widehat{\mathscr{C}_n})\big] < \infty, \quad \text{and}$$

$$\forall U : \mathrm{diam}(U) > b, \quad \mathbb{E}_U\big[\mathrm{diam}(\widehat{\mathscr{C}_n})\big] \leq \mathrm{diam}(U) - n.$$

第一个条件成立，因为根据 (4)， $\mathbb{P}_U$(直径($\mathscr{C}_n$) $\leq$直径(U) $+ nm$) $= 1$。关于第二个条件，注意定理 1.2 适用，因为 $n < 2m + 2$ ，且结合 (4) 可推出

$$\mathbb{E}_U\big[\mathrm{diam}(\widehat{\mathscr{C}_n})\big] \leq (1 - p)(\mathrm{diam}(U) + n) + p(2n^2),$$

对于仅取决于 n 的 $p > 0$。因此，第二个条件成立，其中 $b = (2/p - 1)n + 2n^2$。 □

命题 1.5 （暂态性）。 **</style}如果，那么 在Prog**$_{m\text{中},\ n}$是暂态的。

证明。 根据定理 1.4 和博雷尔–坎泰利引理，若 C 有 $n \geq 2m + 2$ 个元素，则 $\mathscr{C}_t$ 在直径小于 $r > 0$ 的 $\mathsf{Prog}_{m,n}$ 集合中只会访问有限次；对每个 r，这一结论 $\mathbb{P}_C$几乎必然成立，因此 $\mathscr{C}_t$ 是暂态的。 □

定理 1.5 综合了命题 1.1 至 1.5。

定理 1.5 的证明。 根据标准马尔可夫链理论，命题 1.1 至 1.3 蕴含了$\widehat{\ }$ $\widehat{\ }$的唯一平稳分布的存在性，该分布支撑于 $\mathrm{Prog}_{m,\,n}$，并且从任意构型$\widehat{\ }$出发都将收敛于该分布，当 $n < 2m+2$时。命题 1.1 和 1.5 蕴含，当 $n \geq 2m+2$时，$\mathscr{C}_t$ 是暂态的。□

定理 1.5 体现了证明临界基数存在性的一种高层次策略——也就是说，展示一个马尔可夫链，使其随着初始状态的数量增加而从常返转变为暂态。马尔可夫链二分定理（例如，[FMM95，定理 1.2。1]）说明了这一策略的优点。该定理指出，如果马尔可夫链 $\mathscr{X}$ 在可数集 $\mathcal{T}$ 上不可约，那么 $\mathcal{T}$ 中的每个状态要么几乎必然是常返的，要么每个状态都几乎必然是暂态的。这意味着，一个无界函数 $f\colon\ \mathcal{T} \to \mathbb{R}_{\geq 0}$ 具有有限次水平集 $\{\{T \in \mathcal{T}\colon\ f(T) \leq i\}\}_{i\in\mathbb{N}}$ 满足

$$\liminf_{t\geq 0} f(\mathscr{X}_t) \overset{\text{a.s.}}{=} \begin{cases} \inf_{T\in\mathcal{T}} f(T) < \infty & \mathscr{X} \text{ is recurrent,} \\ \infty & \mathscr{X} \text{ is transient.} \end{cases} \tag{13}$$

特别地，马尔可夫链二分定理意味着

$$\forall T \in \mathcal{T},\quad \mathbb{P}_T\left(f(\mathscr{X}_t) \to \infty\right) \in \{0,1\}.$$

现在，假设 $\mathscr{X}$ 具有不可约分量 $\{\mathcal{T}_n\}_{n\geq 2}$，其中 $\mathcal{T}_n$ 的每个元素都是一个 n 元素集。为简单起见，假设 $\mathcal{T} = \cup_{n\geq 2}\mathcal{T}_n$。如果存在 $n_c \in \mathbb{Z}_{\geq 2}$，使得当 $n < n_c$时，$\mathscr{X}$ 在 $\mathcal{T}_n$ 上常返，而当 $n \geq n_c$时，$\mathscr{X}$ 在 $\mathcal{T}_n$ 上暂态，那么由 (13) 可知，$\{f(\mathscr{X}_t) \to \infty\}$ 的临界基数为n_c。

这一观察与定理 1.1 相关，因为 diam 是定义在 $\widehat{\mathcal{S}} = \cup_{n>m}\{\widehat{S}$ 上的无界函数，且具有有限次水平集：$S \subset_n \mathbb{Z}^d\}$，即视平移等价的构型。根据命题 1.1，$\widehat{\ }$ $\widehat{\ }$ $\widehat{\ }$ $\widehat{\ }\mathscr{C}$ 的不可约分量为 $\{\mathrm{Prog}_{m,\,n}\}_{n>m}$，其中每个 $\mathrm{Prog}_{m,\,n}$ 都是 $\mathcal{S}_n = \{S\colon\ S \subset_n \mathbb{Z}^d\}$ 的子集，而 $\widehat{\mathcal{S}}_n$ 的其他元素均为非本质的。根据 (13)，如果 $n > m$，那么

$$\forall \widehat{S} \in \widehat{\mathcal{S}}_n,\quad \mathbb{P}_{\widehat{S}}\big(\mathrm{diam}(\widehat{\mathscr{C}_t}) \to \infty\big) = \begin{cases} 0 & \widehat{\mathscr{C}_t} \text{ is recurrent on } \widehat{\mathrm{P}}\mathrm{rog}_{m,n}, \\ 1 & \widehat{\mathscr{C}_t} \text{ is transient on } \widehat{\mathrm{P}}\mathrm{rog}_{m,n}. \end{cases}$$

$\widehat{\ }$对于 $\mathbb{P}_S(\mathrm{diam}(\mathscr{C}_t) \to \infty)$ 也同样成立，因为 CAT 在平移下保持不变，并且对于每个构型 S，都有 $\mathrm{diam}(S) = \mathrm{diam}(S)$。因此，定理 1.1 可由定理 1.5 推出。

1.5. 结构安排。 第 2 节汇集了 CAT 转移概率的初步估计，为第 3 节讨论两个关键证明思路作准备。在第 4 节中，我们运用第 3 节中的第一个思路，迅速证明定理 1.2。第 5 节以证明定理 1.4 开篇，证明中运用了第 3 节中的第二个思路，并假设一个关键命题成立。第 3 节的其余部分用于证明这一命题，该命题验证了这样一种启发式观点：能够持续存在且彼此充分分离的簇会像随机游走一样移动。接着，第 6 节分两个阶段证明定理 1.3，其中第一阶段再次运用了$\widehat{\ }$第 3 节中的第一个思路。最后一节，即第 7 节，证明 CAT 在 $\mathrm{Prog}_{m,\,n}$上不可约，从而完成定理 1.5 的证明。

1.6. 致谢。 感谢 Milind Hegde、Joseph Slote 和于斌给予的鼓励性讨论；感谢 Sarah Cannon、Joshua Daymude、Brian Daniels、Stephanie Forrest、Dana Randall、Sidney Redner 和 Andréa Richa 提供的宝贵反馈；并感谢 Judit Zádor 和 Noah Egan 建议在表 1 中加入第四个和第六个例子。

2. 预备知识

本节汇总了转移概率的基本估计，以便讨论第 3 节中定理 1.2–1.4 证明所依据的思想。这些估计主要来自直径（4）的几乎必然界，它控制元素之间的距离，并由此控制 CAT 的转移概率。

2.1. CAT 转移概率的公式。 用 $\mathcal{A}_{t,j}$ 和 $\mathcal{T}_{t,j}$ 分别表示在时刻 t被激活和传输的 j^{th} 元素。我们根据上下文，用 $\mathcal{A}_t$ 和 $\mathcal{T}_t$ 表示 $(\mathcal{A}_{t,j})_{j\le m}$ 和 $(\mathcal{T}_{t,j})_{j\le m}$，或 $\{\mathcal{A}_{t,j}\}_{j\le m}$ 和$\{\mathcal{T}_{t,j}\}_{j\le m}$。此外，我们用 $\mathcal{X}_t$ 表示 $\mathcal{A}_t$ 和 $\mathcal{T}_t$的串接，即 $(\mathcal{A}_{t,1},\dots,\mathcal{A}_{t,m},\mathcal{T}_{t,1},\dots,\mathcal{T}_{t,m})$。据此，

$$\mathscr{C}_{t,l}=\begin{cases}\mathscr{C}_t\setminus\{\mathcal{A}_{t,j}\}_{j\le l} & l\le m,\\ (\mathscr{C}_t\setminus\mathcal{A}_t)\cup\{\mathcal{T}_{t,j}\}_{j\le l-m} & l>m.\end{cases}$$

特别地，$\mathscr{C}_{t+1}=(\mathscr{C}_t\setminus\mathcal{A}_t)\cup\mathcal{T}_t$。

我们可以用定义在 $\mathbb{Z}^d$ 上的一族概率测度来表示 CAT 转移概率，这族概率测度以有限且非空的$U\subset\mathbb{Z}^d$ 和 $u\in\mathbb{Z}^d$为索引：

$$\mu_{U,u}(v)=\frac{\phi(u,v)}{\sum_{w\in U}\phi(u,w)}\mathbf{1}_U(v). \tag{14}$$

换言之，$\mu_{U,\,u}$ 是定义在 U 上的概率测度，且与 $\phi(u,\,\cdot)$成正比。为简化记号，我们进一步定义$\mu_{U,\,\infty}(v)=|U|^{-1}\mathbf{1}_U(v)$。用这些测度表示，转移概率的各分量为

$$\mathbb{P}\left(\mathcal{A}_t=x\mid\mathscr{C}_t\right)=\prod_{j=1}^{m}\mu_{(\mathscr{C}_t\setminus\{x_i\}_{i<j}),x_{j-1}}(x_j)\quad\text{and}$$

$$\mathbb{P}\left(\mathcal{T}_t=y\mid\mathscr{C}_t,\mathcal{A}_t\right)=\prod_{j=1}^{m}\mu_{\partial((\mathscr{C}_t\setminus\mathcal{A}_t)\cup\{y_i\}_{i<j}),y_{j-1}}(y_j),$$

对于 $x,\ y\in(\mathbb{Z}^d)^m$，其中 $x_0=\infty$ 和$y_0=\mathcal{A}_{t,\,m}$。CAT 转移概率为

$$\mathbb{P}(\mathscr{C}_{t+1}=D\mid\mathscr{C}_t=C)=\sum_{(x,y):\,C\to D}\mathbb{P}\left(\mathcal{A}_t=x,\mathcal{T}_t=y\mid\mathscr{C}_t\right),$$

对于构型 $C,\ D$，其中 $C\to D$ 表示求和范围为满足以下条件的 $x,\ y\in(\mathbb{Z}^d)^m$

$$D=(C\setminus\{x_i\}_{i\le m})\cup\{y_i\}_{i\le m}.$$

2.2. 预备引理。 现在我们使用几乎必然条件 (4) 以及 $\phi(u,\ v)$ (3) 和$\mu_{U,\,u}(v)$ (14) 的定义，证明 CAT 转移概率的三个基本估计。

引理 2.1。 设 C_0、$E\subset_n\mathbb{Z}^d$、$x\in(\mathbb{Z}^d)^{2m}$和 $t\ge1$，并根据以下定义 $\{C_l\}_{1\le l\le2m}$

$$C_l=\begin{cases}C_0\setminus\{x_j\}_{j\le l} & l\le m,\\ \left(C_0\setminus\{x_j\}_{j\le m}\right)\cup\{x_j\}_{j=m+1}^{l} & l>m.\end{cases} \tag{15}$$

存在 a、$b>0$，使得如果以下概率为正，则它们实际上满足

$$\mathbb{P}_{C_0}(\mathscr{C}_{0,l+1}=C_{l+1}\mid\mathscr{C}_{0,l}=C_l)\ge a\|x_{l+1}-x_l\|^{-b},$$

$$\mathbb{P}_{C_0}(\mathscr{C}_1=C_{2m})\ge a\operatorname{diam}(\{x_j\}_{1\le j\le2m})^{-b},\quad\textit{and}$$

$$\mathbb{P}_{C_0}(\mathscr{C}_t=E)\ge a^t(t\operatorname{diam}(C_0))^{-bt}.$$

证明。 我们逐一考察各个界。首先，根据 CAT 动态的定义 s,

$$\mathbb{P}_{C_0}(\mathscr{C}_{0,l+1}=C_{l+1}\mid \mathscr{C}_{0,l}=C_l)=\begin{cases}\mu_{C_l,x_l}(x_{l+1}) & l\le m\\ \mu_{\partial C_l,x_l}(x_{l+1}) & l>m.\end{cases} \tag{16}$$

根据 (3) 和 (14)，若 $U\subset\mathbb{Z}^d$ 有限且非空，且若 u， $v\in\mathbb{Z}^d$，那么要么 $\mu_{U,\ u}(v)=0$，要么 $\mu_{U,\ u}(v)\ \geq|U|^{-1}\|u-v\|^{-\beta}$。因此，若 (16) 中的条件概率为正，则它满足

$$\mathbb{P}_{C_0}(\mathscr{C}_{0,l+1}=C_{l+1}\mid \mathscr{C}_{0,l}=C_l)\geq(2dn)^{-1}\|x_{l+1}-x_l\|^{-\beta}, \tag{17}$$

因为 $|C_l|\le n$ 且 $|\partial C_l|\le 2dn$。

其次，假设 $\mathbb{P}_{C_0}(\mathscr{C}_1=C_{2m})>0$，此时必定存在满足 (15) 的 $x\in(\mathbb{Z}^d)^{2m}$ 和 $\{C_l\}_{1\le l\le 2m}$，以及满足条件的 $\mathbb{P}_{C_0}(\mathscr{C}_{0,\ 1}=C_1,\ \ldots,\ \mathscr{C}_{0,\ 2m}=C_{2m})>0$。我们有

$$\mathbb{P}_{C_0}(\mathscr{C}_1=C_{2m})\geq\prod_{1\le l\le 2m-1}\mathbb{P}_{C_0}(\mathscr{C}_{0,l+1}=C_l\mid\mathscr{C}_{0,l}=C_l)\geq(2dn)^{-2m}\prod_{1\le l\le 2m-1}\|x_{l+1}-x_l\|^{-\beta}. \tag{18}$$

马尔可夫性质蕴含第一个不等式，也蕴含对每个 $l\in[\![1,\ 2m-1]\!]$，$\mathbb{P}_{C_0}(\mathscr{C}_{0,\ l+1}=C_{l+1}\mid\mathscr{C}_{0,\ l}=C_l)>0$ 这一事实。因此，第二个不等式由 (17) 得出。用 $\mathrm{diam}(\{x_j\}_{1\le j\le 2m})$ 对 $\|x_{l+1}-x_l\|$ 作上界估计，可得

$$\mathbb{P}_{C_0}(\mathscr{C}_1=C_{2m})\geq(2dn)^{-2m}\mathrm{diam}(\{x_j\}_{1\le j\le 2m})^{-2m\beta}. \tag{19}$$

第三，假设 $\mathbb{P}_{C_0}(\mathscr{C}_t=C')>0$，此时存在一个构型序列 $D_0,\ldots,D_t$，满足$D_0=C_0$和 $D_t=E$，且有$\mathbb{P}_{D_0}(\mathscr{C}_1=D_1,\ldots,\mathscr{C}_t=D_t)>0$。考虑 $s\in[\![1,t]\!]$。注意，如果 $y\in(\mathbb{Z}^d)^{2m}$ 和 $\{D_{s,l}\}_{1\le l\le 2m}$ 满足（15），其中分别以 x 和 $\{C_l\}_{1\le l\le 2m}$代替，并且如果 $\mathbb{P}_{D_{s-1}}(\mathscr{C}_{0,1}=D_{s,1},\ldots,\mathscr{C}_{0,2m}=D_{s,2m})>0$，那么几乎必然性质（4）蕴含

$$\|y_{l+1}-y_l\|\le\mathrm{diam}(C_0)+tm\le 2mt\mathrm{diam}(C_0).$$

因此，由（18）可得，

$$\mathbb{P}_{D_{s-1}}(\mathscr{C}_1=D_s)\geq(2dn(2m)^\beta)^{-2m}\big(t\,\mathrm{diam}(C_0)\big)^{-2m\beta}.$$

对每个这样的 s应用马尔可夫性质，可得

$$\mathbb{P}_{C_0}(\mathscr{C}_t=E)\geq(2dn(2m)^\beta)^{-2m}\big(t\,\mathrm{diam}(C_0)\big)^{-2m\beta t}. \tag{20}$$

由（17）、（19）和（20）可得所述的三个界，其中 $a=(2dn(2m)^\beta)^{-2m}$ 且 $b=2m\beta$。 □

概率测度 $\mu_{A,\ u}$ 会按与 $\phi(u,\ v)$ 成比例的方式“选取” $v\in A$。如果 v 实际上属于 $B\subset A$，那么$\mu_{B,\ u}(v)\ \geq\mu_{A,\ u}(v)$，因为 B 中的元素较少。下面的结果进一步表明，$\mu_{A,\ u}(v)$ 至少是 $\mu_{B,\ u}(v)$ 的一个分数；当 $B\cup\{u\}$ 的直径相对于从 u 到 $A\setminus B$的距离更小时，该分数更接近于 1。证明使用了关于 ϕ的简单界。

引理 2.2（比率估计）。 设 $B\subseteq A\subset_n\mathbb{Z}^d$、 $u\in A^c$ 和 $v\in B$。则，

$$\frac{\mu_{A,u}(v)}{\mu_{B,u}(v)}\geq 1-n\frac{\mathrm{diam}(B\cup\{u\})^\beta}{\mathrm{dist}(u,A\setminus B)^\beta}. \tag{21}$$

证明。 根据定义（14），

$$\mu_{A,u}(v)=\frac{\phi(u,v)}{\sum_{z\in B}\phi(u,z)+\sum_{z\in A\setminus B}\phi(u,z)}\quad\text{and}\quad\mu_{B,u}(v)=\frac{\phi(u,v)}{\sum_{z\in B}\phi(u,z)}.$$

由于当$\frac{1}{1+r}\geq 1-r\ \ r\geq 0$时，

$$\frac{\mu_{A,u}(v)}{\mu_{B,u}(v)}\geq 1-\frac{\sum_{z\in A\setminus B}\phi(u,z)}{\sum_{z\in B}\phi(u,z)}.$$

根据假设，$x \notin A$，因此对于每个 $z \in A$，$\phi(u,z) = \|u-z\|^{-\beta}$。由此，

$$\sum_{z\in A\setminus B} \phi(u,z) \le n\mathrm{dist}(x, A\setminus B)^{-\beta} \quad \text{and} \quad \sum_{z\in B}\phi(u,z) \ge \mathrm{diam}(B\cup\{u\})^{-\beta}.$$

所声称的界（21）成立。 □

按照特定序列 $x \in (\mathbb{Z}^d)^{2m}$ 发生激活和传输的概率，是 $|C_0|^{-1}$（用于计入在 x_1处的初始激活）与形如$\mu_{C_l,\ x_l}(x_{l+1})$ 和 $\mu_{\partial C_l,\ x_l}(x_{l+1})$ (16) 的概率测度的乘积。引理 2.2 表明，如果我们将一个集合 D 加入 C_l，那么只要 D 与 C_l 相距较远，且 x_{l+1} 属于 C_l 或 ∂C_l，后面的这些分量就几乎不会改变。下一个结果阐述了这一点；其证明是引理 2.2 的简单应用。

引理 2.3. 设 $C_0 \in \mathcal{S}$，设 D 是 $\mathbb{Z}^d$ 的有限子集，满足 $\mathrm{dist}(C_0, D) \ge 2m$，并设 $x \in (\mathbb{Z}^d)^{2m}$ 满足 $\mathbb{P}_{C_0}(\mathcal{X}_0 = x) > 0$。存在 $c > 0$，使得

$$\frac{\mathbb{P}_{C_0\cup D}(\mathcal{X}_0 = x)}{\mathbb{P}_{C_0}(\mathcal{X}_0 = x)} \ge \frac{|C_0|}{|C_0\cup D|}\left(1 - c|C_0\cup D|\frac{\mathrm{diam}(C_0)^\beta}{\mathrm{dist}(C_0,D)^\beta}\right)^{2m}. \tag{22}$$

证明。 根据 CAT 动力学的定义，

$$\frac{\mathbb{P}_{C_0\cup D}(\mathcal{X}_0 = x)}{\mathbb{P}_{C_0}(\mathcal{X}_0 = x)} = \frac{|C_0|}{|C_0\cup D|}\prod_{l=1}^{m-1}\frac{\mu_{C_l\cup D,x_l}(x_{l+1})}{\mu_{C_l,x_l}(x_{l+1})}\prod_{l=m}^{2m-1}\frac{\mu_{\partial(C_l\cup D),x_l}(x_{l+1})}{\mu_{\partial C_l,x_l}(x_{l+1})}, \tag{23}$$

其中 $(C_l)_{1\le l\le 2m}$ 由 (15) 定义。我们旨在对

$$A = \begin{cases} C_l\cup D & 1\le l<m,\\ \partial(C_l\cup D) & m\le l<2m,\end{cases} \quad B = \begin{cases} C_l & 1\le l<m,\\ \partial C_l & m\le l<2m,\end{cases} \quad u = x_l, \quad \text{and} \quad v = x_{l+1},$$

每个 l 应用引理 2.2。引理 2.2 的假设要求 $B \subseteq A$、$u \in A^c$ 和 $v \in B$。第三项要求由 (15) 满足。前两项要求之所以满足，是因为

$$C_l \subseteq C_l\cup D, \quad \partial C_l \subseteq \partial(C_l\cup D), \quad \text{and} \quad x_l \in (C_l\cup D)^c.$$

事实上，第二点源于假设 $\mathrm{dist}(C_0, D) \ge 2m$，这意味着 D 与 C_l 不相交；第三点也源于这一假设，它意味着 $x_l \in D^c$，以及假设 $\mathbb{P}_{C_0}(\mathcal{X}_0 = x) > 0$，后者意味着 $x_l \in C_l^c$。

根据引理 2.2，对每个 $1 \le l < m$，

$$\frac{\mu_{C_l\cup D,x_l}(x_{l+1})}{\mu_{C_l,x_l}(x_{l+1})} \ge 1 - |C_l\cup D|\frac{\mathrm{diam}(C_l\cup\{x_l\})^\beta}{\mathrm{dist}(x_l,D)^\beta} \ge 1 - |C_0\cup D|\frac{\mathrm{diam}(C_0)^\beta}{\mathrm{dist}(C_0,D)^\beta}.$$

第二个不等式成立，是因为对每个这样的 l，$C_l \subseteq C_0$、$x_l \in C_0$ 和 $C_l\cup\{x_l\} \subseteq C_0$。再次根据引理 2.2，对每个 $m \le l < 2m$，

$$\frac{\mu_{\partial(C_l\cup D),x_l}(x_{l+1})}{\mu_{\partial C_l,x_l}(x_{l+1})} \ge 1 - |\partial(C_l\cup D)|\frac{\mathrm{diam}(\partial C_l\cup\{x_l\})^\beta}{\mathrm{dist}(x_l,D)^\beta} \ge 1 - c|C_0\cup D|\frac{\mathrm{diam}(C_0)^\beta}{\mathrm{dist}(C_0,D)^\beta}.$$

第二个不等式成立，因为对于每个这样的 l，都有 $c = 2d(2m+6)^\beta$

$$\begin{aligned} |\partial(C_l\cup D)| &\overset{*}{=} |\partial C_l| + |D| \le 2d|C_l| + |D| \le 2d|C_0| + |D| \overset{*}{\le} 2d|C_0\cup D|,\\ \mathrm{diam}(\partial C_l\cup\{x_l\}) &\le \mathrm{diam}(\overline{C}_{2m}) \overset{**}{\le} \mathrm{diam}(C_0) + m + 2 \le (m+3)\mathrm{diam}(C_0),\\ \mathrm{dist}(x_l, D) &\overset{**}{\ge} \mathrm{dist}(C_0,D) - m \overset{*}{\ge} \mathrm{dist}(C_0,D)/2. \end{aligned}$$

注意，标有一个星号的等式和不等式由 dist(C_0, D) $\geq 2m$ 这一假设以及几乎必然性成立。标有两个星号的不等式由几乎必然性 (4) 成立。将这些比值界代入 (23) 可得 (22)。□

3. 证明思路

3.1. 思路 #1：获得与直径无关的界的方法。定理 1.2 和定理 1.3 的关键特点是，其结论与初始状态的直径无关，尽管 CAT 转移概率取决于元素之间的距离。这些结果背后的简单思路是：当 CAT 从一个不理想的子集开始激活时，通常会形成一个更“理想”的构型。例如，如果理想的构型是将元素集中在一个“较小”的球内，或由“较大”的连通分量构成，那么这一点便成立。由于激活以均匀随机的方式开始，这一思路便能得出与直径无关的界。

3.1.1. 定理 *1.2* 的应用。定理 1.2 旨在将构型中的元素集中到一个球内，该球的直径不依赖于初始状态的直径。为获得更理想的构型，我们将球外的元素清除，也就是说，令以下事件发生

$$\mathsf{AddToBall}_{x,r} := \big\{|\mathscr{C}_1 \cap \mathcal{B}_x(r+m)| > |\mathscr{C}_0 \cap \mathcal{B}_x(r)|\big\},$$

其定义于 $x \in \mathbb{Z}^d$ 和 $r \in \mathbb{R}_{>0}$。证明定理 1.2 的关键在于给出该事件发生概率的一个下界。

引理 3.1. 若 $n \leq 2m+1$ 且 $C \subset_n \mathbb{Z}^d$ 满足 $|C \cap \mathcal{B}_x(r)| \in [\![m+1, n-1]\!]$ 则对于某些 $x \in \mathbb{Z}^d$ 和 $r \geq 1$，有

$$\mathbb{P}_C(\mathsf{AddToBall}_{x,r}) \geq p, \tag{24}$$

其中 $p > 0$ 仅依赖于 r。

证明。 思路 *#1* 的表述是

$$\mathbb{P}_C(\mathsf{AddToBall}_{x,r} \mid J > 1) \geq p, \tag{25}$$

对于某些仅依赖于 r 的 $p > 0$，有 $J := \inf\{j : \mathcal{A}_{0,\,j} \in \mathcal{B}_x(r)\}$。要证明 (24)，只需改为证明 (25)，因为

$$\mathbb{P}_C(J > 1) = \mathbb{P}_C(\mathcal{A}_{0,1} \notin \mathcal{B}_x(r)) \geq \frac{1}{n}.$$

该不等式由以下事实得出：根据假设，$\mathcal{A}_{0,1}$ 在 C 中均匀随机，其中至少有一个元素位于 $\mathcal{B}_x(r)$之外。

我们声称

$$\mathsf{AddToBall}_{x,r} \cap \{J > 1\} \supseteq \mathsf{AddToBall}' \cap \{J > 1\}, \tag{26}$$

其中，添加到球′ 是事件

$$\{J = \infty\} \cup \{1 < J \leq m,\ \{\mathcal{A}_{0,j}\}_{j=J}^{m} \subset \mathcal{B}_x(r),\ \mathcal{T}_0 \subseteq \mathcal{B}_x(r+m)\}.$$

换句话说，要么所有激活都发生在 $\mathcal{B}_x(r)$之外，要么一旦 $\mathcal{B}_x(r)$中的某个元素被激活，之后所有激活都留在 $\mathcal{B}_x(r)$内，传输也留在 $\mathcal{B}_x(r+m)$内。我们可以在不估计 $\mathbb{P}_C(J = \infty \mid J > 1)$的情况下，给出添加到球′ 发生概率的下界。

已知 $\{J > 1\}$ 发生，添加到球′ 以正概率发生，因为根据假设，$\mathcal{B}_x(r)$包含超过m 个来自 C 的元素。于是引理 2.1意味着存在 $a,\ b > 0$ 使得

$$\mathbb{P}_C(\mathsf{AddToBall}' \mid J > 1) \geq a r^{-b}$$

因为，当添加到球′ 发生时，($\mathcal{A}_{0,J}$、……、$\mathcal{A}_{0,m}$、$\mathcal{T}_{0,1}$、……、$\mathcal{T}_{0,m}$) 中相邻元素之间的距离至多为 $2(r+m) \lesssim r$。这在满足 (26) 的条件下证明了 (25)，其中 $p = a r^{-b}$。

包含关系 (26) 成立，因为如果 $\{J=\infty\}$ 发生，那么 $\mathcal{B}_x(r)$ 之外的每个元素都会被激活；这些元素必然会被传输到 $\mathcal{T}_0 \subseteq \mathcal{B}_x(r+m)$。或者，如果添加到球$'\cap\{1<J\le m\}$发生，那么添加到球$_{x,\ r}$ 就会发生，因为 $\mathcal{B}_x(r)$ 之外至少有一个元素被激活，并且每个被激活的元素都会被传输到 $\mathcal{B}_x(r+m)$。□

3.1.2. 定理 *1.3* 的应用。定理 1.3 旨在构造一种构型，其中包含彼此充分分离且由 $m+1$ 到 $2m+1$ 个元素组成的线段。此过程的第一阶段构造一种由连通分量组成的构型，这些连通分量在每个都含有多于 m 个元素的意义上是"较大的"。为了得到更理想的构型，我们移除小分量中的元素，也就是说，通过事件的发生

$$\mathsf{DepleteSmallComps}_t = \{|\mathscr{R}_{t+1}| < |\mathscr{R}_t|\}, \tag{27}$$

该事件针对 $t\in\mathbb{Z}_{\ge 0}$ 定义，其形式为 $\mathscr{R}_t := \{x\in\mathscr{C}_t:\ |\mathrm{comp}_{\mathscr{C}_t}(x)|\le m\}$，其中 $\mathrm{comp}_C(y)$ 表示在 $C\subset\mathbb{Z}^d$ 中与 $y\in\mathbb{Z}^d$ 连通的元素集合。定理 1.3 的证明所需的关键输入，是该事件发生概率的一个下界。

引理3.2。存在 $p>0$，使得如果 $C\subset_n\mathbb{Z}^d$ 包含某个 $x\in\mathbb{Z}^d$，且 $|\mathrm{comp}_C(x)|\le m$，那么

$$\mathbb{P}_C(\mathsf{DepleteSmallComps}_0)\ge p. \tag{28}$$

证明。 思路 *#1* 的表述是以下断言：

$$\mathbb{P}_C(\mathsf{DepleteSmallComps}_0 \mid K>1)\ge p, \tag{29}$$

对于某个 $p>0$，其中 re $K := \inf\{k:\mathcal{A}_{0,\ k}\in\mathscr{C}_0\setminus\mathscr{R}_0\}$。要证明 (28)，只需证明 (29)，因为 e

$$\mathbb{P}_C(K>1)=\mathbb{P}_C(\mathcal{A}_{0,1}\in\mathscr{R}_0)\ge\frac{1}{n}.$$

该不等式由以下事实推出：根据假设，$\mathcal{A}_{0,\ 1}$ 在 C 中均匀随机，其中至少有一个元素属于 $\mathscr{R}_0$。

我们声称

$$\mathsf{DepleteSmallComps}_0\cap\{K>1\}\supseteq\mathsf{DepleteSmallComps}'\cap\{K>1\}, \tag{30}$$

其中，耗尽小连通分量′ 是以下事件

$$\Big(\{K=\infty\}\cup\big\{1<K\le m,\ \{\mathcal{A}_{0,k}\}_{k=K}^{m}\cup\{\mathcal{T}_{0,1}\}\subseteq\mathrm{comp}_{\mathscr{C}_0}(\mathcal{A}_{0,K})\big\}\Big)\cap\{\mathcal{T}_0\text{ is connected in }\mathscr{C}_1\}.$$

换言之，括号中的事件发生，当且仅当激活只发生在小型连通分量中；或者，一旦大型连通分量中的某个元素被激活，之后所有激活都发生在该分量中，以及传输首次到达的地点。

给定 $\{K>1\}$ 发生，耗尽小连通分量′ 以正概率发生，因为根据假设，$\mathscr{C}_0\setminus\mathscr{R}_0$ 的每个连通分量都包含超过 m 个元素。因此，引理 2.1 表明，存在$p>0$ 使得

$$\mathbb{P}_C(\mathsf{DepleteSmallComps}'\mid K>1)\ge p$$

因为，当耗尽小连通分量′ 发生时，序列 ($\mathcal{A}_{0,\ K}$，……，$\mathcal{A}_{0,\ m}$，$\mathcal{T}_{0,\ 1}$，……，$\mathcal{T}_{0,\ m}$) 中相邻元素之间的距离至多为 n。这证明了在满足 (30) 的条件下，(29) 成立。

包含关系 (30) 成立，因为如果发生 $\{K=\infty,\mathcal{T}_0$ 在 $\mathscr{C}_1\}$ 中连通，那么每个 被激活的元素都会新加入一个大型连通分量，或与其被传输到其边界的 $\mathscr{C}_0\setminus\mathcal{A}_0$ 中的元素共同形成一个大型连通分量$_{,\ 1}$。如果发生的是 $\mathsf{DepleteSmallComps}'\cap\{1<K\le m\}$，那么 $\mathsf{DepleteSmallComps}_0$ 也会发生，因为 $\{\mathcal{A}_{0,k}\}_{k<K}$ 处的元素会新加入大型连通分量 $\mathrm{comp}_{\mathscr{C}_0}(\mathcal{A}_{0,\ K})$，其余被激活的元素则返回该分量。□

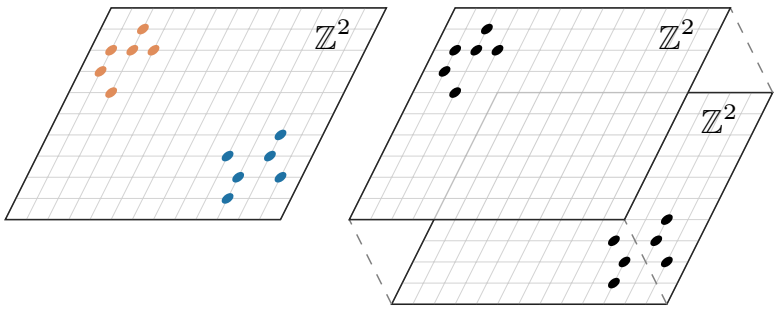

图5。CAT 的提升对应于一种不受动力学影响的着色（左），而 CopyCAT 构型中的元素可能位于 $\mathbb{Z}^2$ 的不同“副本”中（右）。

3.2. 思想2：对 CAT 动力学日益精确的近似。定理 1.4 的证明的关键思想是，当由 $n \geq 2m+2$ 个元素构成的构型包含至少 $m+1$ 个元素的簇时，CAT 动力学的行为近似于这些簇分别占据 $\mathbb{Z}^d$ 的不同“副本”。这种近似动力学比 CAT 动力学更容易分析。至关重要的是，随着簇间分离度增大，这种近似的准确性也会提高。因此，我们可以用它来估计那些意味着分离增长足够快的事件的概率，包括定理 1.4 中的事件。

为精确说明这一点，我们将 CAT“提升”为状态空间 Part 上的马尔可夫链 $\mathscr{C}_t^I = (\mathscr{C}_t^i)_{i\in I}$。Part 是 $\mathcal{S}$ 中构型的有序划分所构成的集合，每个分划都包含多于 m 个元素。这里，I 是形如 $[\![1,\ k]\!]$ 的区间，其中 $k \in \mathbb{Z}_{\geq 1}$。给定 $\mathscr{C}_t^I$，构造 $\mathscr{C}_{t+1}^I$ 时，我们只需将 CAT 动力学应用于 $\mathscr{C}_t := \cup_{i\in I}\mathscr{C}_t^i \in \mathcal{S}$，并更新划分，以反映元素的新位置。换言之，我们将划分视为元素的固定着色，它在动力学中不起作用。需要强调的是，在划分 C^I 的语境下，我们用 C 表示 $\cup_{i\in I}C^i$。

接下来，我们将 *CopyCAT* 定义为状态空间为配置元组 Tup $= \cup_{i\geq 1}\mathcal{S}^i$ 的马尔可夫链 $\mathscr{D}_t^I = (\mathscr{D}_t^i)_{i\in I}$，其动力学如下。给定 $\mathscr{D}_t^I$，要构造 $\mathscr{D}_{t+1}^I$，我们将 CAT 动力学应用于一个条目 $\mathscr{D}_t^{J_t}$，该条目按照与 $|\mathscr{D}_t^i|$ 成正比的概率随机选取。换言之，

$$\mathscr{D}_{t+1}^i := \begin{cases} \mathrm{CAT}(\mathscr{D}_t^i) & i = J_t, \\ \mathscr{D}_t^i & i \neq J_t, \end{cases} \tag{31}$$

其中，对于任意 $C,\ C' \in \mathcal{S}$ 和 $i \in I$,

$$\mathbb{P}(\mathrm{CAT}(C) = C') = \mathbb{P}_C(\mathscr{C}_1 = C') \quad \text{and} \quad \mathbb{P}(J_t = i \mid \mathscr{D}_t) = \frac{|\mathscr{D}_t^i|}{\sum_{i\in I}|\mathscr{D}_t^i|}.$$

我们用 $\mathcal{Y}_t$ 表示在时刻 t 被激活并传输的 $\mathscr{D}_t^I$ 的元素序列，即 CopyCAT 对应于 $\mathcal{X}_t$的类比。

为陈述关键近似结果，我们将 $U^I \in$ Tup 的分离定义为

$$\mathrm{sep}(U^I) := \min_{i\in I} \mathrm{dist}(U^i, U^{\neq i}),$$

其中 $U^{\neq i} = \cup_{j\in I\setminus\{i\}}U^j$。对于 $a,\ b \in \mathbb{R}_{>0}$，我们将绝对分离度为 a 且相对于直径的 分离度 b 的元组定义为

$$\mathrm{Tup}_{a,b} := \left\{U^I \in \mathrm{Tup} : \mathrm{sep}(U^I) \geq a \quad \text{and} \quad \forall i \in I,\ \mathrm{diam}(U^i) \leq b \log \mathrm{dist}(U^i, U^{\neq i})\right\}. \tag{32}$$

注意，由于 a 为正数，因此 $\mathrm{Tup}_{a,\,b} \subseteq \mathrm{Part}$。对于 $\gamma \in \mathbb{R}_{>0}$，$\delta \in (0,\,1/2)$，以及 $t \in \mathbb{Z}_{\geq 0}$，我们将 $\mathrm{Growth}_{a,\,b,\,\gamma,\,\delta;t}$ 定义为满足以下条件的序列 $(U_s^I)_{s\geq 0} \in \mathrm{Tup}_{a,\,b}^{\mathbb{N}}$ 的集合

$$\forall s \leq t, \quad \mathrm{sep}(U_s^I) \geq \gamma s^{1/2-\delta}.$$

此外，我们定义 $\mathrm{Growth}_{a,\,b,\,\gamma,\,\delta} = \cap_{t\geq 0}\mathrm{Growth}_{a,\,b,\,\gamma,\,\delta;t}$。

命题 3.1 (关键近似结果)。令 $b, \gamma > 0$ 和 $\delta \in (0, \frac{1}{2} - \frac{1}{\beta})$。若 $a \geq 2m$ 足够大，则

$$\mathbb{P}_{C^I}\left((\mathscr{C}_t^I)_{t\geq 0} \in \mathrm{Growth}_{a,b,\gamma,\delta}\right) \geq (1 - o_a(1))\,\mathbb{P}_{C^I}\left((\mathscr{D}_t^I)_{t\geq 0} \in \mathrm{Growth}_{a,b,\gamma,\delta}\right)$$

对于每个 $C^I \in \mathrm{Tup}_{a,\,b}$。

命题 3.1 的证明主要基于对 CAT 和 CopyCAT 单步转移概率的比较。

命题 3.2 (单步近似)。令 $b > 0$，并令 $U^I, V^I \in \mathrm{Tup}_{2m,\,b}$ 满足 $\mathbb{P}_{U^I}(\mathscr{C}_1^I = V^I) > 0$。存在 $c > 0$ 使得

$$\mathbb{P}_{U^I}\left(\mathscr{C}_1^I = V^I\right) \geq \left(1 - c|U|\,\mathrm{sep}(U^I)^{-\beta}\big(\log \mathrm{sep}(U^I)\big)^{\beta}\right)^{2m} \mathbb{P}_{U^I}\left(\mathscr{D}_1^I = V^I\right). \tag{33}$$

命题 3.1 是马尔可夫性质、命题 3.2 以及 $\mathrm{Growth}_{a,\,b,\,\gamma,\,\delta}$ 的定义的简单推论。

Pro 命题 *3.1* 的令 $t \geq 0$。由马尔可夫性质和命题 3.2，我们有 e

$$\begin{aligned}\mathbb{P}_{C^I}\left((\mathscr{C}_s^I)_{s\geq 0} \in \mathrm{Growth}_{a,b,\gamma,\delta;t}\right) &= \sum \mathbb{P}_{C^I}\left(\mathscr{C}_s^I = D_s^I, s \geq 0\right) \\ &\geq w_t \sum \mathbb{P}_{C^I}\left(\mathscr{D}_s^I = D_s^I, s \geq 0\right) = w_t \mathbb{P}_{C^I}\left((\mathscr{D}_s^I)_{s\geq 0} \in \mathrm{Growth}_{a,b,\gamma,\delta;t}\right),\end{aligned}$$

其中求和遍及 $(D_s^I)_{s\geq 0} \in \mathrm{Growth}_{a,\,b,\,\gamma,\,\delta;t}$。这里，$w_t = \prod_{s\leq t}\left(1 - cr_a(s)^{-\beta}(\log\, r_a(s))^{\beta}\right)^{2m}$，其中 $c > 0$ 是（33）中同名的数，而 $r_a(s) = \max\{a, \gamma s^{1/2-\delta}\}$。由于 $\mathrm{Growth}_{a,\,b,\,\gamma,\,\delta;t}$ 随 t 递减，$\mathbb{P}_{C^I}$ 的连续性意味着

$$\mathbb{P}_{C^I}\left((\mathscr{C}_s^I)_{s\geq 0} \in \mathrm{Growth}_{a,b,\gamma,\delta}\right) \geq \lim_{t\to\infty} w_t\, \mathbb{P}_{C^I}\left((\mathscr{D}_s^I)_{s\geq 0} \in \mathrm{Growth}_{a,b,\gamma,\delta}\right).$$

容易看出，由于 $\beta > 2$，若 $\delta < \frac{1}{2} - \frac{1}{\beta}$，则 $\sum_{s\geq 0} r_a(s)^{-\beta}(\log\, r_a(s))^{\beta} = o_a(1)$，这意味着 $\lim_{t\to\infty} w_t = (1 - o_a(1))$。□

命题 3.2 是引理 2.3 的一个应用。

命题 *3.2* 的证明。固定 U^I，$V^I \in \mathrm{Tup}_{2m,\,b}$ 并定义 $\mathcal{G}$ 和 $\mathcal{H}$ 分别为激活序列和传输序列 $x \in (\mathbb{Z}^d)^{2m}$，它们在 CAT 和 CopyCAT 动力学下由 U^I 构成 V^I：

$$\mathcal{G} := \left\{x \in (\mathbb{Z}^d)^{2m} : \{\mathscr{C}_0^I = U^I, \mathcal{X}_0 = x\} \subseteq \{\mathscr{C}_1^I = V^I\} \text{ and } \mathbb{P}_{U^I}(\mathcal{X}_0 = x) > 0\right\},$$

$$\mathcal{H} := \left\{x \in (\mathbb{Z}^d)^{2m} : \{\mathscr{D}_0^I = U^I, \mathcal{Y}_0 = x\} \subseteq \{\mathscr{D}_1^I = V^I\} \text{ and } \mathbb{P}_{U^I}(\mathcal{Y}_0 = x) > 0\right\}.$$

请注意，$\mathcal{G} \supseteq \mathcal{H}$ 因为 V^I 彼此间的距离超过了 m。事实上，只要这些簇之间的距离过大，以至于在 CopyCAT 动力学运行一步后仍不会相交，$\mathcal{G}$ 与 $\mathcal{H}$ 之间的唯一区别就在于存在 $x \in \mathcal{G} \setminus \mathcal{H}$，它们对应于在多个簇上的激活。

我们计算

$$\mathbb{P}_{U^I}(\mathscr{C}_1^I = V^I) = \sum_{x\in\mathcal{G}} \mathbb{P}_{U^I}(\mathcal{X}_0 = x) \geq \sum_{x\in\mathcal{H}} \mathbb{P}_{U^I}(\mathcal{X}_0 = x) \geq \sum_{x\in\mathcal{H}} w_x \frac{|U^{i_x}|}{|U|} \mathbb{P}_{U^{i_x}}(\mathcal{X}_0 = x), \tag{34}$$

其中， i_x 是 U^I 的簇， x_1 属于该簇，并且

$$w_x = \left(1 - c|U|\frac{\operatorname{diam}(U^{i_x})^\beta}{\operatorname{dist}(U^{i_x}, U^{\neq i_x})^\beta}\right)^{2m},$$

根据引理 2.3 中的量 c。等式 (34) 根据定义成立，第一个不等式由$\mathcal{G} \supseteq \mathcal{H}$得出，第二个不等式则是将引理 2.3 应用于 $C_0 = U^{i_x}$和 $D = U^{\neq i_x}$所得。使用该引理要求 dist(C_0, D) $\geq 2m$ 且 $\mathbb{P}_{C_0}$($\mathcal{X}_0 = x$) > 0。前一个条件之所以满足，是因为U^I的分离至少为 $2m$；后一个条件之所以满足，是因为 $\mathbb{P}_{U^I}(\mathcal{Y}_0 = x)$为正，这是由$x \in \mathcal{H}$保证的，而这意味着 $\mathbb{P}_{C_0}(\mathcal{X}_0 = x)$也为正，因为

$$\mathbb{P}_{U^I}(\mathcal{Y}_0 = x) = \frac{|U^{i_x}|}{|U|} \mathbb{P}_{U^{i_x}}(\mathcal{X}_0 = x)$$

由 (31) 可知。根据定义， $\mathbb{P}_{U^I}(\mathscr{D}_1^I = V^I) = \sum_{x\in\mathcal{H}} \mathbb{P}_{U^I}(\mathcal{Y}_0 = x)$，因此将前一个公式代入 (34)，我们得出

$$\mathbb{P}_{U^I}(\mathscr{C}_1^I = V^I) \geq \min_{x\in\mathcal{H}} w_x \, \mathbb{P}_{U^I}(\mathscr{D}_1^I = V^I).$$

所声称的界（33）由下式推出

$$\min_{x\in\mathcal{H}} w_x \geq \min_{j\in I} \left(1 - c|U|\frac{\big(b \log \operatorname{dist}(U^j, U^{\neq j})\big)^\beta}{\operatorname{dist}(U^j, U^{\neq j})^\beta}\right)^{2m} \geq \left(1 - cb^\beta|U|\frac{\big(\log \operatorname{sep}(U^I)\big)^\beta}{\operatorname{sep}(U^I)^\beta}\right)^{2m}.$$

第一个不等式成立，因为 $U^I \in$Tup$_{\cdot,b}$；第二个不等式成立，因为 dist($U^j, U^{\neq j}$) $\geq$sep({v11)，且 $\log_x{}^x$ 关于 $x > e$递减。□

4. 直径坍缩

定理 1.2 指出，若 $n \leq 2m + 1$，则 $\mathscr{C}_n$ 的直径至多为 $2n^2$，且其概率至少为某个 $p > 0$。我们通过估计 τ_x 来证明定理 1.2；对于 $x \in \mathbb{Z}^d$，将其定义为首次满足以下条件的时间 t： $\mathscr{C}_t$ 中位于球 $\mathcal{B}_x(tm + n)$ 内的元素数量 $N_x(t)$ 等于 n:

$$\tau_x := \inf\{t \geq 0 : N_x(t) = n\} \quad \text{where} \quad N_x(t) := |\mathscr{C}_t \cap \mathcal{B}_x(tm + n)|. \tag{35}$$

当 $\{\tau_x \leq m\}$ 发生时， $\mathscr{C}_n$ 的直径满足

$$\operatorname{diam}(\mathscr{C}_n) \leq 2(\tau_x m + n) + (n - \tau_x)m = \tau_x m + n(m + 2) \leq 2n^2. \tag{36}$$

第一个不等式由 τ_x 的定义以及几乎必然成立的 (4) 得出。第二个不等式则通过分别以 m、 $n - 1$ 和 n^2 对 τ_x、 m 和 $2n^2 - n + 1$ 进行界定而得出。

4.1. 引理 3.1 的应用。 下一个命题指出，对于某些 $C \subset_n \mathbb{Z}^d$， $\{\tau_x \leq m\}$ 通常会在 $\mathbb{P}_C$ 下发生。该证明是引理 3.1 的一个应用。在 (36) 的背景下，这就证明了此类 C 对应的定理 1.2。

命题 4.1. 设 $C \subset_n \mathbb{Z}^d$ 满足 $|C \cap \mathcal{B}_x(n)| \geq m + 1$，其中某个 $x \in \mathbb{Z}^d$。若 $n \in [\![m + 1, 2m + 1]\!]$，则存在 $p > 0$ 使得

$$\mathbb{P}_C(\tau_x \leq m) \geq p. \tag{37}$$

证明。对于 $0 \le j < k \le m$，记 $A_{j,k}$ 为事件，即 $N_x(t)$从 $t = j$ 到$t = k$的每一步都递增，即 $A_{j,k} = \{N_x(j) < \cdots < N_x(k)\}$。为方便记号，定义 $A_{0,0}$ 为样本空间。关键观察是

$$\mathbb{P}_C(\tau_x \le m \mid \tau_x > 0) \ge \mathbb{P}_C(A_{0,\tau_x} \mid \tau_x > 0). \tag{38}$$

事实上，根据假设，$\mathcal{B}_x(n)$至少包含 C 中的 $n \le 2m+1$ 个元素中的 $m+1$ 个元素，因此 $N_x(t)$连续增长超过 $n-(m+1)$次的可能性不大。

引理 3.1 意味着（38）式的右端至少为 q^m，其中某个 $q > 0$满足这一条件，因此（37）式以 $p = q^m$成立。为说明这一点，根据以下关系定义 $(f_j,\ g_j,\ r_j)_{1\le j\le m}$

$$\begin{aligned} f_j &:= \mathbb{P}_C(A_{j-1,\tau_x} \mid \tau_x > j-1, A_{0,j-1}),\\ g_j &:= \mathbb{P}_C(A_{j-1,j} \mid \tau_x > j-1, A_{0,j-1}),\\ r_j &:= \mathbb{P}_C(\tau_x > j \mid \tau_x > j-1, A_{0,j-1}), \end{aligned}$$

并注意到 f_1 是（38）式的右端。很容易验证，这些量满足以下方程组：

$$f_m = g_m \quad \text{and} \quad \forall 1 \le j \le m-1, \quad f_j = g_j(1 - r_j + r_j f_{j+1}).$$

特别地，对于每个 $j \le m-1$，无论r_j的值如何，$f_j \ge g_j f_{j+1}$ 都成立。因此，要证明 $f_1 \ge q^m$，只需证明对于每个 $j \in [\![1,\ m]\!]$，$g_j \ge q$ 成立。给定 $\mathscr{C}_j$，当 $\{\tau_x > j-1, A_{0,j-1}\}$发生时，$N_x(j)$属于 $[\![m+1,\ n-1]\!]$，因此 $\mathscr{C}_j$ 满足引理 3.1陈述中对 C 的假设。于是，在 $t = j$处应用马尔可夫性质后，引理 3.1意味着，对于某个 $q > 0$，$g_j \ge q$ 成立。□

4.2. 定理 1.2 的证明。我们旨在应用命题 4.1，但其假设规定，C 必须允许一个 $x \in \mathbb{Z}^d$，使得 $|C \cap \mathcal{B}_x(n)| \ge m+1$。为满足这一假设，我们对 $t = 1$ 应用马尔可夫性质，并在 $G = \{\mathcal{T}_0 \subset \mathcal{B}_{\mathcal{T}_{0,1}}(n)\}$ 发生时，用 $\mathscr{C}_1$ 代替 C。（当 G 发生时，$\mathscr{C}_1$满足$|\mathscr{C}_1 \cap \mathcal{B}_{\mathcal{T}_{0,1}}(n)| \ge m+1$。）由引理 2.1 可知，存在 $q_1 > 0$ 使得 $\mathbb{P}_C(G) \ge q_1$，因此只需证明

$$\mathbb{P}_{\mathscr{C}_1}(\operatorname{diam}(\mathscr{C}_{n-1}) \le 2n^2) \ge q_2 \mathbf{1}_G, \tag{39}$$

存在某个 $q_2 > 0$，因为这样(6)就会在 $p = q_1 q_2$下成立。

给定 $\mathscr{C}_1$，当 G 发生时，令 $x \in \mathbb{Z}^d$ 满足 $|\mathscr{C}_1 \cap \mathcal{B}_x(n)| \ge m+1$，并按照(35)定义 $N_x(t)$和 τ_x。由命题 4.1，存在 $q_2 > 0$ 使得

$$\mathbb{P}_{\mathscr{C}_1}(\tau_x \le m) \ge q_2 \mathbf{1}_G. \tag{40}$$

这个界意味着（39），因为当 $\{\tau \le m\}$ 出现时，我们有

$$\operatorname{diam}(\mathscr{C}_{n-1}) \le 2(\tau_x m + n) + (n - 1 - \tau_x)m \le 2n^2.$$

与（36）一样，第一个不等式由 τ_x 的定义和（4）得出，而第二个不等式成立是因为 $\tau_x \le m \le n-2$ 和 $2n^2 - 5n + 6 \le 2n^2$。

5. 直径增长

5.1. 定理 1.4 的证明。回忆一下，对于 $a,\ b,\ \gamma > 0$和 $\delta \in (0,\ 1/2)$，集合 $\mathsf{Growth}_{a,\ b,\ \gamma,\ \delta}$由满足以下条件的 $(C_t^I)_{t\ge 0} \in \mathsf{Tup}^{\mathbb{N}}_{a,\ b}$ 组成

$$\forall t \ge 0, \quad \operatorname{sep}(C_t^I) \ge \gamma t^{\frac{1}{2}-\delta}.$$

本节将证明，只要初始划分充分分离，CopyCAT 通常就会表现出这种分离增长。

命题 5.1。设 $C \subset_n \mathbb{Z}^d$ 对于 $n \geq 2m+2$。存在 $b, \gamma > 0$，使得如果 $C \in \mathsf{CanPersist}_a$ 对于充分大的 $a \in \mathbb{Z}_{\geq 0}$成立，那么对于每个 $\delta \in (0, \frac{1}{2})$，

$$\mathbb{P}_{C^I}\left((\mathscr{D}_t^I)_{t\geq 0} \in \mathsf{Growth}_{\tilde{a},b,\gamma,\delta}\right) \geq \frac{1}{4},$$

其中，$\tilde{a} = a^{0.99}$ 和 C^I 是将 C 划分为满足(7)的 a-分离线段的任意一种划分。[4]

我们用 $a^{0.99}$ 代替 a，因为分离度暂时降至低于 a 的某个常数倍 a是常见情况；当 a足够大时，我们可以将其吸收到 $a^{0.01}$ 中。

定理 1.4 可由　　定理 1.3、关键近似结果（命题 3.1）以及命题 5.1 推出　　.

定理 *1.4* 的证明。令 $C \subset_n \mathbb{Z}^d$ 取 $n \geq 2m+2$，令 b 和 γ 分别为命题 5.1 中同名的数，令 $\delta \in (0, \frac{1}{2} - \frac{1}{\beta})$，并令 $a \in \mathbb{Z}_{\geq 2m}$、$\tilde{a} = a^{0.99}$和 $s = 2n^2 a$。由马尔可夫性质应用于时刻 s可得，

$$\mathbb{P}_C\left(\operatorname{diam}(\mathscr{C}_t) \geq \gamma(t-s)^{\frac{1}{2}-\delta} \text{ for every } t \geq s\right) \geq \mathbb{E}_C\left[\mathbb{P}_{\mathscr{C}_s^I}\left((\mathscr{C}_t^I)_{t\geq 0} \in \mathsf{Growth}_{\tilde{a},b,\gamma,\delta}\right) \mathbf{1}_{\{\mathscr{C}_s \in \mathsf{CanPersist}_a\}}\right],$$

其中，给定 $\mathscr{C}_s \in \mathsf{CanPersist}_a$，$\mathscr{C}_s^I$是满足 (7) 的 $\mathscr{C}_s$ 的任意划分。由命题 3.1 和 5.1 可知，若 a 足够大，则

$$\mathbb{P}_{D^I}\left((\mathscr{C}_t^I)_{t\geq 0} \in \operatorname{Growth}_{\tilde{a},b,\gamma,\delta}\right) \geq \frac{1}{2}\mathbb{P}_{D^I}\left((\mathscr{D}_t^I)_{t\geq 0} \in \operatorname{Growth}_{\tilde{a},b,\gamma,\delta}\right) \geq \frac{1}{8}, \tag{41}$$

对于满足 (7) 的 $D \in \mathsf{CanPersist}_a$ 的每个划分 D^I。注意，我们需要 $D^I \in \mathsf{Tup}_{\tilde{a},\, b}$，才能用命题 3.1 证明第一个不等式。当 $\tilde{a} > e^{n/b}$ 时，这一点成立，因为根据 (7)，对于 $\tilde{b} = {}^{n}\!\log_{\tilde{a}}$，有 $D^I \in \mathsf{Tup}_{\tilde{a},\, \tilde{b}}$，而根据 (32)，有 $\mathsf{Tup}_{\tilde{a},\, \tilde{b}} \subseteq \mathsf{Tup}_{\tilde{a},\, b}$。

根据前述界和定理 1.3，存在 $p > 0$ 使得

$$\mathbb{P}_C\left(\operatorname{diam}(\mathscr{C}_t) \geq \gamma(t-s)^{\frac{1}{2}-\delta},\ t \geq s\right) \geq \frac{1}{8}\mathbb{P}_C(\mathscr{C}_s \in \mathsf{CanPersist}_a) \geq \frac{1}{8}p^a.$$

虽然我们假设 $\delta < \frac{1}{2} - \frac{1}{\beta}$ 以应用命题 3.1，但这个界意味着 (9) 对每个 $\delta \in (0, \frac{1}{2})$ 都成立。□

–

在本节余下部分，我们将通过分析随机游走来证明命题 5.1。这些随机游走由两个簇中代表性元素之差构成，元素依照 CopyCAT 动力学演化，并在一系列更新时刻进行观察。随机游走的标准估计表明，如果这些代表之间的分离足够大，那么随着初始分离的增大，其分离将以高概率按 $\Omega(t^{1/2-\delta})$ 增长。

5.2. **与成对簇相关的随机游走**。给定 $\mathscr{D}_t^I$，对于每个 $i \in I$ 和 $t \geq 0$，我们用 M_t^i 表示 $\mathscr{D}_t^i$，其定义为 $\mathscr{D}_t^i$ 中按照 $\mathbb{Z}^d$ 上的字典序排列最靠前的元素。[5]我们将在所有簇都构成平行于 e_1 的线段（即 $\mathcal{L}$ 中的集合）的连续时刻，考察每个 M_t^i。将这些时刻定义为以下序列：

$$\xi_0 := \inf\left\{t \geq 0 : \mathscr{D}_t^i \in \mathcal{L} \text{ for every } i \in I\right\}, \quad \text{and}$$
$$\forall l \geq 1, \quad \xi_l := \inf\left\{t > \xi_{l-1} : \mathscr{D}_t^i \in \mathcal{L} \text{ for every } i \in I\right\}.$$

[4]注意到，C 必定存在这样的划分，因为 $C \in \mathsf{CanPersist}_a$。[5]例如，用质心来代表簇并无益处，因为簇的直径相对于它们之间的分离很小。因此，元素将同等地代表其所属的簇；与质心不同的是，它们必然属于 $\mathbb{Z}^d$，这会很方便。

我们将分别研究每一对不同的簇 $i,\ j \in I$，通过随机游走 $(S_k^{ij})_{k\geq 0}$，其定义如下:

$$S_k^{ij} := \left(M_{\xi_0}^i - M_{\xi_0}^j\right) + \sum_{l=1}^{k} \left(M_{\xi_l}^i - M_{\xi_l}^j - M_{\xi_{l-1}}^i + M_{\xi_{l-1}}^j\right). \tag{42}$$

部分是为了 确保 $\xi_0 \equiv 0$，我们将在 $\mathbb{P}_{D^I}$ 条件下研究 $(S_k^{ij})_{k\geq 0}$，对于满足该条件的元组 $D^I \in \mathrm{Tup}$ y

$$|I| \geq 2 \quad \text{and} \quad \forall i \in I, \quad D^i \in \mathcal{L} \quad \text{and} \quad |D^i| \in [\![m+1, 2m+1]\!]. \tag{43}$$

命题 5.2。 若 $D^I \in \mathrm{Tup}$ 满足 (43)，那么，对于任意不同的 i、$j \in I$，在 $\mathbb{P}_{D^I}$ 条件下，$(S_k^{ij})_{k\geq 0}$ 的分布是 $\mathbb{Z}^d$上的对称、非周期且不可约的随机游走。

证明。 首先， S_k^{ij} 是对称随机游走，因为 CopyCAT 的相变概率是平移不变的。其次，它是非周期的，因为对于每个 $t \geq 0$， $\mathscr{D}_t^I$的每个簇都属于 Prog_m, $\mathbb{P}_{D^I}$– 几乎必然成立。因此，根据渐进边界性质 (10)，我们有 $\mathbb{P}_{D^I}(\mathscr{D}_1 = \mathscr{D}_0) > 0$，这意味着$\mathbb{P}_{D^I}(S_1^{ij} = S_0^{ij}) > 0$。第三，关于不可约性，只需证明

$$\forall l \in [\![1, d]\!], \quad \mathbb{P}_{D^I}\left(\xi_1 = 5, \mathscr{D}_5^i = \mathscr{D}_0^i + e_l, \mathscr{D}_5^j = \mathscr{D}_0^j\right) > 0,$$

因为相关事件的发生意味着 $S_1^{ij} = S_0^{ij} + e_l$。我们可以通过以下方式实现这一事件。令 $l \in [\![1, d]\!]$，并且不失一般性地假设 $\mathscr{D}_0^i = L_{1,n_i}$，其中 $n_i = |\mathscr{D}_0^i|$。第一步，我们要求 $\mathscr{D}_1^j \notin \mathcal{L}$，以免在 $t = 5$之前意外到达 ξ_1。然后，在接下来的两步中，我们将 $\mathscr{D}_1^i = \mathscr{D}_0^i$变为 $\{e_1\} \cup (L_{1,n_i-1} + e_l)$，这在假设 $n_i \leq 2m+1$的情况下是可行的。第四步我们到达$\mathscr{D}_4^i = L_{1,n_i} + e_l = \mathscr{D}_0^i + e_l$，第五步则反向执行第一步。这一系列步数得到 $\xi_1 = 5$ 和 $S_1^{ij} = S_0^{ij} + e_l$。 □

下一个结果表明，只要相邻两步之间的时间相对较短，与一对簇相关联的随机游走的范数就可以作为它们分离程度的替代指标。我们将在下一节分析相邻两次更新之间的时间。

命题 5.3. 设 $l \geq 1$，设 $t \in [\![\xi_{l-1}, \xi_l]\!]$，并设 $D^I \in \mathrm{Tup}$ 满足(43)。那么，

$$\mathrm{dist}(\mathscr{D}_t^i, \mathscr{D}_t^j) \geq \left\|S_{l-1}^{ij}\right\| - O\left((\xi_l - \xi_{l-1})^2\right) \tag{44}$$

对每一对不同的 $i, j \in I$，几乎必然有 $\mathbb{P}_{D^I}$。

证明。 我们将证明

$$\left\|(M_t^i - M_t^j) - (M_{\xi_{l-1}}^i - M_{\xi_{l-1}}^j)\right\| \lesssim (\xi_l - \xi_{l-1})^2. \tag{45}$$

注意到$\|S_{l-1}^{ij}\| = \left\|M_{\xi_{l-1}}^i - M_{\xi_{l-1}}^j\right\|$，由三角不等式、(4) 和 (45) 即可得到所声称的界 (44):

$$\begin{aligned}
\mathrm{dist}(\mathscr{D}_t^i, \mathscr{D}_t^j) &\geq \|M_t^i - M_t^j\| - \mathrm{diam}(\mathscr{D}_t^i) - \mathrm{diam}(\mathscr{D}_t^j) \\
&\geq \|M_t^i - M_t^j\| - 2(\xi_l - \xi_{l-1}) - |D^i| - |D^j| \\
&\geq \left\|M_{\xi_{l-1}}^i - M_{\xi_{l-1}}^j\right\| - O\left((\xi_l - \xi_{l-1})^2\right).
\end{aligned}$$

为证明（45），我们先将差写成望远镜求和，并应用三角不等式

$$\left\|(M_t^i - M_t^j) - (M_{\xi_{l-1}}^i - M_{\xi_{l-1}}^j)\right\| \leq \sum_{s=\xi_{l-1}}^{t-1} \left(\left\|M_{s+1}^i - M_s^i\right\| + \left\|M_{s+1}^j - M_s^j\right\|\right).$$

容易看出， $\|M^i_{s+1} - M^i_s\|$ 至多为 $\operatorname{diam}(\mathscr{D}^i_s) + m$，因此由（4）可得

$$\|M^i_{s+1} - M^i_s\| \le (\xi_l - \xi_{l-1}) + \operatorname{diam}(\mathscr{D}^i_{\xi_{l-1}}) + m,$$

对于 $s \in [\![\xi_{l-1},\ \xi_l]\!]$。由于 $\mathscr{D}^i_{\xi_{l-1}}$ 是长度为 $|D^i|$ 的线段，其直径为 $|D^i|$，并且根据 $\xi_l - \xi_{l-1} \ge 1$，我们可以简单地用 $(|D^i| + m + 1)(\xi_l - \xi_{l-1})$ 来界定 $\|M^i_{s+1} - M^i_s\|$。将此应用于前面的求和式，可得

$$\left\|(M^i_t - M^j_t) - (M^i_{\xi_{l-1}} - M^j_{\xi_{l-1}})\right\| \le (|D^i| + |D^j| + 2m + 2)(\xi_l - \xi_{l-1})^2.$$

□

5.3. **更新时刻估计**。本小节的主要结果表明， $\xi_l - \xi_{l-1}$ 通常会以 $\log l$ 的速度增长，具体而言：

命题 5.4。存在 α_1、 $\alpha_2 \in \mathbb{R}_{>0}$，使得如果 $D^I \in \mathrm{Tup}$ 满足 (43)，那么

$$\mathbb{P}_{D^I}\left(\cap_{l=1}^{\infty} F_l\right) \ge \frac{1}{2}, \tag{46}$$

其中 $F_l :\!= \{\xi_l - \xi_{l-1} \le \alpha_1 \log(\alpha_2 l)\}$。

证明源于这样一个事实： $\xi_1 - \xi_0$ 是指数紧的。

命题 5.5。 存在 c_1、 $c_2 \in \mathbb{R}_{>0}$，使得若 $D^I \in \mathrm{Tup}$ 满足(43)，则

$$\forall t \ge 0, \quad \mathbb{P}_{D^I}(\xi_1 > t) \le c_1 e^{-c_2 t}. \tag{47}$$

命题 *5.4* 的证明。 对 ξ_{l-1} 应用强马尔可夫性质，并结合 (47)，可得

$$\mathbb{P}_{D^I}(F_l^c) \le c_1 e^{-c_2 \alpha_1 \log(\alpha_2 l)}.$$

取 $\alpha_1 = \frac{2}{c_2}$ 和 $\alpha_2 = \frac{2}{\sqrt{c_1}}$，得到一个上界 $(2l)^{-2}$。再应用并集界即可得到 (46)　　：

$$\mathbb{P}_{D^I}\left(\cap_{l=1}^{\infty} F_l\right) \ge 1 - \sum_{l=1}^{\infty} \mathbb{P}_{D^I}(F_l^c) \ge 1 - \sum_{l=1}^{\infty} (2l)^{-2} \ge \frac{1}{2}.$$

□

连续两次更新之间的时间具有指数紧性，原因有二。首先，令 $D^I \in$ 为满足 $\sum_{i \in I} |D^i| = n$ 的 Tup。根据 CopyCAT 的定义 (31)，

$$\mathbb{P}_{D^I}\left((\mathscr{D}^i_s)_{s \le t} = (C_s)_{s \le t}\right) \ge n^{-t}\, \mathbb{P}_{D^i}\left((\mathscr{C}_s)_{s \le t} = (C_s)_{s \le t}\right), \tag{48}$$

对每个 $i \in I$、 $t \ge 1$，以及 $(C_s)_{s \le t} \in \mathcal{S}^t$ 都成立。其次，在 $\mathbb{P}_C$ 下，CAT 经过固定步数后返回 $\mathcal{L}$ 的概率至少为某个 $q > 0$；对于满足 $n \in [\![m+1, 2m+1]\!]$ 个元素的所有 $C \subset_n \mathbb{Z}^d$，这一结论一致成立。

命题 5.6。 存在一个正数 $p > 0$，使得如果 $C \subset_n \mathbb{Z}^d$ 有 $n \in [\![m+1,\ 2m+1]\!]$ 个元素，那么

$$\mathbb{P}_C(\mathscr{C}_{n+2} \in \mathcal{L}) \ge p. \tag{49}$$

(49) 中的时间 $n+2$ 来自等待 n 步，使直径足够小（即由 n 的某个函数所界定）（定理 1.2）；之后，我们可以指定两个传输步骤来形成一条线，而不会使 p 依赖于 $\operatorname{diam}(C)$。（我们很快会解释如何用两步形成一条线。）

命题 *5.5* 的证明。设 $E^I \in$ Tup 满足 $|E^i| \in [\![m+1, 2m+1]\!]$ 对每个 $i \in I$ 均成立，并记作 $s = \sum_{i\in I}(|E^i|+2)$。由 (48) 和命题 5.6，存在 $p > 0$ 使得

$$\mathbb{P}_{E^I}\left(\mathscr{D}_s^i \in \mathcal{L} \text{ for all } i \in I\right) \geq p,$$

对于任意这样的 E^I。注意到 $s \leq 2n$，因为 E^I 至多有 $\lfloor \frac{n}{m+1} \rfloor \leq \frac{n}{2}$ 个簇，且 $\sum_{i\in I}|E^i| = n$。因此，如果 $D^I \in$ Tup 满足 (43)，那么

$$\forall t \geq 0, \quad \mathbb{P}_{D^I}(\xi_1 > t) \leq (1-p)^{\lfloor \frac{t}{s} \rfloor} \leq (1-p)^{\frac{t}{2n}-1}.$$

第一个不等式是通过应用前述界以及每隔 s 步应用马尔可夫性质得到的。第二个不等式利用了以下事实：对于 $r > 0$ 和 $s \leq 2n$，$\lfloor r \rfloor \geq r - 1$。这就以 $c_1 = \frac{1}{1-p}$ 和 $c_2 = \frac{1}{2n}\log(\frac{1}{1-p})$. 证明了 (47)。 □

作为本小节的总结，我们证明命题 5.6。证明利用了这样一个事实：如果 C 有 $n \leq 2m+1$ 个元素，那么 $\mathscr{C}_2$ 形成一条直线，且其概率至少为一个取决于 C 直径的值。

命题 5.7。 存在正数 c_1、c_2，使得如果 $C \subset_n \mathbb{Z}^d$ 有 $n \in [\![m+1$ 个 $2m+1]\!]$ 元素，那么

$$\mathbb{P}_C(\mathscr{C}_2 \in \mathcal{L}) \geq c_1 \mathrm{diam}(C)^{-c_2}.$$

事实上，对于任意 $n > m$，若以 $\mathscr{C}_{n-1}$ 替代 $\mathscr{C}_2$，且 $n-1$ 是 m 的倍数，则同样的界也成立；否则需要 $m\mathscr{C}_{\lfloor \frac{n-1}{m} \rfloor + 1}$。

证明。设 $x \in C$ 是 C 中 e_1 分量最大的元素。第一步，激活除 x 外任意 m 个元素，并将它们传输到 $x+e_1$、……、$x+me_1$。第二步，激活剩余的 $n-(m+1)$ 个 $\leq m$ 元素，以及第一步中激活的 $m-(n-(m+1))$ 个元素。将前一类元素放置在 $x-e_1$、……、$x-(n-(m+1))e_1$，后一类元素则保持原位。当这些步骤发生时，

$$\mathscr{C}_2 = \{x-(n-(m+1))e_1, \ldots, x+me_1\} \in \mathcal{L}.$$

CAT 以正概率通过这种方式实现 $\mathcal{L}$。由引理 2.1 即可得到所声称的界。我们省略对 $n > m$ 的推广，因为这很直接。□

命题 5.6 是定理 1.2 和命题 5.7 的直接推论。

命题 *5.6* 的证明。记

$$\mathbb{P}_C(\mathscr{C}_{n+2} \in \mathcal{L}) \geq \mathbb{P}_C(\mathrm{diam}(\mathscr{C}_n) \leq 2n^2)\, \mathbb{P}_C(\mathscr{C}_{n+2} \in \mathcal{L} \mid \mathrm{diam}(\mathscr{C}_n) \leq 2n^2).$$

由定理 1.2，第一项至少为 $q_1 > 0$；而由马尔可夫性质，在时刻 n 应用命题 5.7，可知第二项至少为 $q_2 > 0$。取 $p = q_1 q_2$ 即可证明结论。□

5.4. **随机游走范数的增长**。 在本小节中，我们考虑任意一对不同的簇。为简单起见，将这两个簇标记为 1 和 2，并从所有记号中省略 ij 下标。例如，我们将 S_k^{12}记为 S_k。在本小节中，始终假设 $D^I \in$ Tup 满足 (43)；此时，根据命题 5.2，$(S_k)_{k\geq 0}$ 在 $\mathbb{P}_{D^I}$ 下是 $\mathbb{Z}^d$上的对称、非周期、不可约随机游走。

我们需要用到两个标准的随机游走估计。记 $(S_k)_{k\geq 0}$ 进入 $A \subseteq \mathbb{Z}^d$ 的首次时刻为 $T_A := \inf\{k \geq 0 : S_k \in A\}$。 [LL10] 的命题 2.4.5 和 6.4.2 指出，存在 $\alpha_3, \alpha_4, \alpha_5 \in \mathbb{R}_{>0}$，使得

对于任意 $R, \lambda \in \mathbb{R}_{>0}$ 和 $x \in B(R)^c$,

$$\mathbb{P}_x(T_{B(R)^c} > \lambda R^2) \le \alpha_3 e^{-\alpha_4 \lambda}, \tag{50}$$

$$\mathbb{P}_x(T_{B(R)} < \infty) \le \alpha_5 \left(\frac{R}{\|x\|}\right)^{d-2}. \tag{51}$$

注5.1（尽管增量无界，这些估计仍然成立）。 [LL10] 中的命题*2.4.5*和命题*6.4.2*假设随机游走的增量有界，而 S_k并非如此。不过，由于 S_k 的增量具有指数紧性（见命题 *5.5*），这些命题的证明仍可照原样适用。唯一的例外是，在命题*6.4.2*的证明中，必须将对命题*4.3.1*的引用替换为对命题*4.3.5*的引用，结论相同。

注5.2（对任意一对簇，都可以使用相同的常数）。 一般而言， α_3、 α_4和 α_5 取决于随机游走的增量分布。然而，由 i 和 j所索引的游走，其增量分布至多有（$m+1$）^{2}种，因为每个簇的元素数目都属于 $[\![m+1, 2m+1]\!]$。因此，我们可以选取常数，使（50） 和（51） 对每一对簇都成立。

令 $r > 1$ 并定义

$$w(r) = \left[\alpha_2^{-1} \log\left(2\alpha_1 (\log r)^2\right)(er)^2\right] + 1 \quad \text{and} \quad \delta(r) = \frac{1}{2\alpha_3 (\log r)^2}. \tag{52}$$

命题 5.8。若 $x \in \mathbb{Z}^d$ 满足 $r := \|x\| > 1$，则

$$\mathbb{P}_x\left(T_{B(er)^c} \le w(r),\ T_{B(\delta(r)r)} = \infty\right) \ge 1 - \frac{1}{(\log r)^2}.$$

证明。由并集界，只需证明 $\mathbb{P}_x(T_{B(er)^c} > w(r))$ 和 $\mathbb{P}_x(T_{B(\delta(r)r)} < \infty)$ 均不超过 $\frac{1}{2(\log r)^2}$。将 $w(r)$ 和 $\delta(r)$代入 (50) 和 (51) 即可得到这些界。□

定义随机时刻

$$\begin{aligned} \eta_l(r) &:= \inf\{k \ge \eta_{l-1}(r) : \|S_k\| > e\|S_{\eta_{l-1}(r)}\|\} \\ \theta_l(r) &:= \inf\{k \ge \eta_{l-1}(r) : \|S_k\| \le \delta(S_{\eta_{l-1}(r)})\|S_{\eta_{l-1}(r)}\|\}. \end{aligned} \tag{53}$$

此外，根据下式定义事件序列 $(G_l(r))_{l\ge1}$

$$G_l(r) := \left\{\eta_l(r) - \eta_{l-1}(r) \le w\left(\|S_{\eta_{l-1}(r)}\|\right),\ \theta_l(r) = \infty\right\}.$$

用这些术语来说，命题 5.8 表明

$$\forall x \in \mathbb{Z}^d : \|x\| > 1, \quad \mathbb{P}_x(G_1(\|x\|)) \ge 1 - \frac{1}{(\log\|x\|)^2}.$$

命题 5.9. 设 $x \in \mathbb{Z}^d$ 满足 $r := \|x\| > 1$。则

$$\mathbb{P}_x\left(\cup_{l=1}^{\infty} G_l(r)^c\right) \lesssim \frac{1}{\log r}. \tag{54}$$

证明。 记作 $H_l(r) = \cap_{k\le l} G_k(r)$。我们将证明

$$q_j := \mathbb{P}_x\left(G_{l+1}(r)^c \,\middle|\, H_l(r)\right) \le \frac{1}{(\log(e^l r))^2}, \tag{55}$$

这蕴含式（54），因为

$$\mathbb{P}_x\left(\cup_{l=1}^{\infty} G_l(r)^c\right) \le \sum_{l=0}^{\infty} q_l \lesssim \frac{1}{\log r}.$$

为证明 (55)，我们对由 $(S_0, \ldots\ldots, S_{\eta_l(r)})$ 生成的 σ-域使用塔式法则，并对 $\eta_l(r)$应用强马尔可夫性质，可得

$$\mathbb{P}_x\left(G_{l+1}(r) \mid H_l(r)\right) = \mathbb{E}_x\left[\mathbb{P}_{S_{\eta_l(r)}}\left(G_1\big(\|S_{\eta_l(r)}\|\big)\right) \frac{\mathbf{1}_{H_l(r)}}{\mathbb{P}_x(H_l(r))}\right]. \tag{56}$$

当 $H_l(r)$ 发生时，$\|S_{\eta_l(r)}\| > e^l r$，因此命题 5.8 蕴含

$$\mathbb{P}_{S_{\eta_l(r)}}\left(G_1\big(\|S_{\eta_l(r)}\|\big)\right) \ge 1 - \frac{1}{(\log(e^l r))^2}.$$

将此界代入 (56) 即得 (55)。 □

下一个命题控制一对簇的绝对分离和相对分离度，当

$$\mathcal{F} := \cap_{l\ge 1} F_l \quad \text{and} \quad \mathcal{G}(r) := \cap_{l\ge 1} G_l(r) \tag{57}$$

在 $r > 0$ 足够大时发生。我们用 $R_l := \|S_{\eta_l}\|$ 来表述它，即随机游走 $l \in \mathbb{Z}_{\ge 0}$“加倍”的范数。[6]

命题 5.10（时间、直径与分离界）。设 $l \in \mathbb{Z}_{\ge 0}$。若 R_0 足够大，则 $\mathcal{F} \cap \mathcal{G}(R_0)$ 的出现意味着

$$t \lesssim (R_l \log R_l)^2, \quad \operatorname{diam}(\mathscr{D}_t^1) \lesssim \log R_l, \quad \operatorname{dist}(\mathscr{D}_t^1, \mathscr{D}_t^2) \gtrsim \frac{R_l}{(\log R_l)^2}, \tag{58}$$

对所有 $t \in N^{-1}\big([\eta_l, \eta_{l+1})\big)$。

前两个界是由 $\mathcal{F}$ 和 $\mathcal{G}(r)$ 的定义推得的，而第三个界还需要利用命题 5.3 将簇间分离度与 R_l 联系起来。命题 5.10 的难点在于它涉及三种不同的时间计数：t 计算 $\mathscr{D}_t$ 的步数，k 计算导出的随机游走 S_k 的步数，而 l 计算 $\|S_{\eta_l}\|$ 的倍增次数。为协调这些计数——尤其是为了证明 (58) 的第一个界——我们将分别用 ξ_k 和 η_l 估计 t；用 η_l 估计 ξ_k；以及用 R_l 估计 η_l。在证明中，我们将用记号 $f \lesssim_{R_0} g$ 表示：在假设 R_0 足够大的情况下，$f \lesssim g$ 成立。

命题 *5.10* 的证明。我们可以按以下方式用 ξ_k 和 η_l 给出 t、直径$(\mathscr{D}_t^1)$ 和 $\operatorname{dist}(\mathscr{D}_t^1, \mathscr{D}_t^2)$ 的上界：

$$t \le \sum_{0\le i\le l} \sum_{j=\eta_i}^{\eta_{i+1}-1} (\xi_{j+1} - \xi_j), \tag{59}$$

$$\operatorname{diam}(\mathscr{D}_t^1) \le \xi_{N(t)+1} - \xi_{N(t)}, \quad \text{and} \tag{60}$$

$$\operatorname{dist}(\mathscr{D}_t^1, \mathscr{D}_t^2) \ge \|S_{N(t)}\| - O\left((\xi_{N(t)+1} - \xi_{N(t)})^2\right). \tag{61}$$

第一个界成立，是因为根据假设，到时间 t 时还没有发生 $l+1$ 次翻倍。第二个界成立，是因为 $\mathscr{D}^1_{\xi_{N(t)}}$ 的直径至多为 n，并且在 $t \le \xi_{N(t)+1}$ 之前，它至多随时间线性增长（这是 (4) 和 (31) 的推论）。第三个界是命题 5.3。

[6]回想一下，$\|S_{\eta_l}\|$ 是随机游走首次超过 $e\|S_{\eta_{l-1}}\|$ 时的范数 (53)。特别地，R_l 至少为 $e^l R_0$.

我们利用 $\mathcal{F}\cap\mathcal{G}(R_0)$ 的发生来用 R_l 界定 η_l，并用 η_l 界定 ξ_k。根据定义，当 $\mathcal{F}\cap\mathcal{G}(R_0)$ 发生时，我们有

$$\forall l\geq 0,\quad \eta_{l+1}-\eta_l\leq w(R_l), \tag{62}$$

$$\forall k\geq 0,\quad \xi_{k+1}-\xi_k\leq \alpha_1\log(\alpha_2(k+1)),\quad \text{and} \tag{63}$$

$$\forall k\in[\eta_l,\eta_{l+1}),\quad \|S_k\|\geq\delta(R_l)R_l. \tag{64}$$

由于 $w(s)$ 的增长速度快于 s^2 (52)，不难看出

$$\eta_{l+1}=\sum_{0\leq j\leq l}(\eta_{j+1}-\eta_j)\overset{(62)}{\leq}\sum_{0\leq j\leq l}w(R_j)\lesssim_{R_0} w(R_l). \tag{65}$$

因此，如果 $k\in[\eta_l,\eta_{l+1})$，那么

$$\xi_{k+1}-\xi_k\overset{(63)}{\leq}\alpha_1\log(\alpha_2\eta_{l+1})\overset{(65)}{\lesssim_{R_0}}\log w(R_l). \tag{66}$$

最后，根据定义 (52)，

$$w(R_l)\lesssim_{R_0} R_l^2\log R_l\quad\text{and}\quad\delta(R_l)\gtrsim(\log R_l)^{-2}. \tag{67}$$

鉴于 (59)，前面三个界表明，t 至多是约 $R_l^2\log R_l$ 个加数之和，每个加数都不超过约 $\log R_l$。更确切地说，我们有

$$t\overset{(59)}{\leq}\sum_{0\leq i\leq l}\sum_{j=\eta_i}^{\eta_{i+1}-1}(\xi_{j+1}-\xi_j)\overset{(65),\,(66)}{\lesssim_{R_0}} w(R_l)\log w(R_l)\overset{(67)}{\lesssim_{R_0}}(R_l\log R_l)^2.$$

这就证明了 (58) 中的第一个界。其余两个界由下式得出：

$$\begin{aligned}\operatorname{diam}(\mathscr{D}_t^1)&\overset{(60)}{\leq}\xi_{N(t)+1}-\xi_{N(t)}\overset{(66)}{\lesssim_{R_0}}\log w(R_l)\overset{(67)}{\lesssim_{R_0}}\log R_l,\\ \operatorname{dist}(\mathscr{D}_t^1,\mathscr{D}_t^2)&\overset{(61)}{\geq}\|S_{N(t)}\|-O\left((\xi_{N(t)+1}-\xi_{N(t)})^2\right)\\ &\overset{(64)}{\geq}\delta(R_l)R_l-O\left((\log w(R_l))^2\right)\overset{(67)}{\gtrsim_{R_0}}R_l(\log R_l)^{-2}.\end{aligned}$$

□

5.5. **命题 5.1 的证明**。现在我们考虑所有不同的簇对，并利用上一节的结果证明簇之间的整体分离不断增大，这正是命题 5.1 的含义。因此，我们重新引入 ij 上标，并定义

$$\mathcal{F}=\cap_{i,j\in I:\,i<j}\mathcal{F}^{ij}\quad\text{and}\quad\mathcal{G}(r)=\cap_{i,j\in I:\,i<j}\mathcal{G}^{ij}(r)$$

对于 $r>0$，其中 $\mathcal{F}^{ij}$ 和 $\mathcal{G}^{ij}(r)$ 是式 (57) 中的事件，我们之前分别记为 $\mathcal{F}$ 和 $\mathcal{G}(r)$。此外，我们定义 $R_0=\min_{i,j\in I:\,i<j}R_0^{ij}$，其中 $R_0^{ij}=\|S_0^{ij}\|$。

接下来的两个结果表明，存在 $b,\gamma>0$，使得对于任意 $\delta\in(0,1/2)$，只要 R_0 足够大，就有

$$\{\mathcal{F}\cap\mathcal{G}(R_0)\}\subseteq\left\{(\mathscr{D}_t^I)_{t\geq0}\in\operatorname{Tup}^{\mathbb{N}}_{\tilde{R}_0,b}\right\}\cap\left\{\operatorname{sep}(\mathscr{D}_t^I)\geq\gamma t^{\frac{1}{2}-\delta}\right\}=\left\{(\mathscr{D}_t^I)_{t\geq0}\in\mathsf{Growth}_{\tilde{R}_0,b,\gamma,\delta}\right\}, \tag{68}$$

其中 $\tilde{R}_0=R_0^{0.99}$。我们需要用 $\tilde{R}_0$ 替代 R_0，因为 $\mathcal{G}(R_0)$ 的定义允许分离暂时降至 $\delta(R_0)$ R_0（参见 $\theta_l(r)$（53）和 $G_l(r)$ 的定义）。这两个证明都是命题 5.10 的应用。

命题 5.11。 存在 $b>0$，使得如果 R_0 足够大，那么

$$\{\mathcal{F}\cap\mathcal{G}(R_0)\}\subseteq\left\{(\mathscr{D}_t^I)_{t\geq 0}\in \mathrm{Tup}^{\mathbb{N}}_{\tilde{R}_0,b}\right\},$$

其中 $\tilde{R}_0=R_0^{0.99}$。

证明。 要使 $(\mathscr{D}_t^I)_{t\geq 0}$ 属于 $\mathrm{Tup}^{\mathbb{N}}_{r,\,b}$，(i) $\mathscr{D}_t^I$ 的分离必须始终至少为 r，并且 (ii) $\mathrm{diam}(\mathscr{D}_t^i)\leq b\log \mathrm{dist}(\mathscr{D}_t^i,\mathscr{D}_t^{\neq i})$ 必须对每个 $t\geq 0$ 和 $i\in I$ 成立。根据命题 5.10，如果当 R_0 充分大时发生 $\mathcal{F}\cap\mathcal{G}(R_0)$，那么

$$\mathrm{dist}(\mathscr{D}_t^i,\mathscr{D}_t^{\neq i})\gtrsim R_l(\log R_l)^{-2}\gtrsim R_0(\log R_0)^{-2}\quad\text{and}\quad \mathrm{diam}(\mathscr{D}_t^i)\lesssim\log\mathrm{dist}(\mathscr{D}_t^i,\mathscr{D}_t^{\neq i}).$$

第一个界说明，当 b 足够大时，(i) 成立；而第二个界说明，当 R_0 充分大时，(ii) 成立。□

Prop命题 5.12。 存在 $\gamma>0$，使得对于任意 $\delta\in(0、\,1\text{–}2)$，如果 R_0 充分大，则 n

$$\{\mathcal{F}\cap\mathcal{G}(R_0)\}\subseteq\left\{\mathrm{sep}(\mathscr{D}_t^I)\geq\gamma t^{\frac{1}{2}-\delta}\right\}.$$

证明。 固定一对 $i,\ j\in I$，使得 $i<j$ 和 $l\in\mathbb{Z}_{\geq 0}$。只需证明存在 $\gamma>0$，使得对于任意 $\delta\in($0、$\,1\,2)$，如果 R_0^{ij} 充分大，则 $\mathcal{F}^{ij}\cap\mathcal{G}^{ij}(R_0^{ij})$ 的发生意味着

$$\mathrm{dist}(\mathscr{D}_t^i,\mathscr{D}_t^j)\geq\gamma t^{\frac{1}{2}-\delta},$$

对每个 $t\in N^{-1}\big([\eta_l^{ij},\eta_{l+1}^{ij}]\big)$ 和 $\delta\in(0,\ 1\,2)$。根据命题 5.10，若 R_0^{ij} 足够大，则 $\mathcal{F}^{ij}\cap\mathcal{G}^{ij}(R_0^{ij})$ 的发生意味着

$$t\lesssim(R_l^{ij}\log R_l^{ij})^2\quad\text{and}\quad\mathrm{dist}(\mathscr{D}_t^i,\mathscr{D}_t^j)\gtrsim R_l^{ij}(\log R_l^{ij})^{-2}.\tag{69}$$

因此，

$$\frac{\mathrm{dist}(\mathscr{D}_t^i,\mathscr{D}_t^j)}{t^{\frac{1}{2}-\delta}}\gtrsim\frac{(R_l^{ij})^{2\delta}}{(\log R_l^{ij})^{3-2\delta}}\geq\frac{(R_0^{ij})^{2\delta}}{(\log R_0^{ij})^{3-2\delta}}.$$

若 R_0^{ij} 相对于 δ 足够大，从而满足 $R_0^{2\delta}\geq(\log R_0)^{3-2\delta}$，则前述界表明 $\mathrm{dist}(\mathscr{D}_t^i,\ \mathscr{D}_t^j)\gtrsim t^{1/2-\delta}$。□

前两个命题确立了包含关系 (68)。要证明命题 5.1，还需说明当 R_0 足够大时，$\mathcal{F}\cap\mathcal{G}(R_0)$ 通常会发生。该证明结合命题 5.4 和命题 5.9，并使用并集界。

命题 5.13。设 $D^I\in\mathrm{Tup}$ 满足 (43)。若 R_0 足够大，则

$$\mathbb{P}_{D^I}(\mathcal{F}\cap\mathcal{G}(R_0))\geq\frac{1}{4}.$$

证明。 首先，命题 5.9 指出，对于满足 $i<j$ 的任意一对簇 $i,\ j\in I$，事件 $\mathcal{G}^{ij}(R_0^{ij})$ 发生的概率至少为 $1-O\big((\log R_0^{ij})^{-1}\big)$。对至多 n^2 对簇应用并集界可知，事件 $\mathcal{G}(R_0)$ 发生的概率至少为 $1-O\big(n^2(\log R_0)^{-1}\big)$。假设 R_0 足够大，使得这个界小于 $3\,4$。其次，根据命题 5.4，若 D^I 满足（43），则 $\mathbb{P}_{D^I}(\mathcal{F})\geq\frac{1}{2}$。这两个界共同推出 $\mathbb{P}_{D^I}(\mathcal{F}\cap\mathcal{G}(R_0))\geq\frac{1}{4}>0$。□

命题 *5.1* 的证明。 设 $a \in \mathbb{Z}_{\geq 0}$ 和 $C \in \mathsf{CanPersist}_a$。根据定义（7），存在一个将 C 划分为 $D^I \in$ 个 Tup 的划分，该划分满足（43），且与之关联的随机游走起初相距 $R_0 \geq a$。因此，根据命题5.11和5.12，存在 $b,\ \gamma > 0$，使得对于任意 $\delta \in$（0，1 2），如果 a 足够大，那么

$$\{\mathcal{F} \cap \mathcal{G}(R_0)\} \overset{(68)}{\subseteq} \left\{(\mathscr{D}_t^I)_{t\geq 0} \in \mathsf{Growth}_{\tilde{R}_0,b,\gamma,\delta}\right\} \subseteq \left\{(\mathscr{D}_t^I)_{t\geq 0} \in \mathsf{Growth}_{\tilde{a},b,\gamma,\delta}\right\},$$

其中 $\tilde{R}_0 = R_0^{0.99}$和 $\tilde{a} = a^{0.99}$。所声称的界由命题 5.13 推出。 □

6. 簇形成

本节证明定理 1.3，该定理涉及能够持久存在的彼此充分分离的簇的形成（7）。证明分两步进行。首先，我们使用引理 3.2 将元素归入包含多于 m 个元素的连通分量。其次，我们将这些连通分量划分为若干部分，每部分包含介于 $m+1$ 和 $2m+1$ 个元素之间，然后逐一将它们“像跑步机一样”分开。

6.1. **停留引理**。 我们需要一个关于 CAT 的基本事实。如果 $U \subset_n \mathbb{Z}^d$ 满足

$$\exists V = \{v_1, \dots, v_m\} \subset_m U: \quad \operatorname{diam}(V) \leq n \quad \text{and} \quad v_l \in \partial\big(U \setminus \{v_l, \dots, v_m\}\big),\ 1 \leq l \leq m, \tag{70}$$

那么 CAT 通常会“停留”在 U，因为它可以激活位于 $v_1, \dots, v_m$ 的元素，然后将它们直接送回。为陈述下一个结果，对于由 $\mathbb{Z}^d$ 的 n元素子集组成的集合 $\mathcal{U}$，定义 $T_{\mathcal{U}} := \inf\{t \geq 0 : \mathscr{C}_t \in \mathcal{U}\}$。

引理 6.1 (停留引理)。 设 $\mathcal{U}$ 是由 $\mathbb{Z}^d$ 的 n 元子集组成的集合。若每个 $U \in \mathcal{U}$ 都满足(70)，且存在 $t \in \mathbb{Z}_{\geq 0}$ 和 $q > 0$，使得对每个 $C \subset_n \mathbb{Z}^d$ 都有 $\mathbb{P}_C(T_{\mathcal{U}} \leq t) \geq q$，则存在 $p > 0$ 使得

$$\mathbb{P}_C(\mathscr{C}_t \in \mathcal{U}) \geq p^t$$

对每个 $C \subset_n \mathbb{Z}^d$ 都成立。

证明。 由于 $\mathbb{P}_C(T_{\mathcal{U}} \leq t) \geq q$ 且 $\{\mathscr{C}_t \in \mathcal{U}, T_{\mathcal{U}} \leq t\} \subseteq \{\mathscr{C}_{T_{\mathcal{U}}} = \cdots = \mathscr{C}_t, T_{\mathcal{U}} \leq t\}$，我们有

$$\mathbb{P}_C(\mathscr{C}_t \in \mathcal{U}) \geq q\mathbb{P}_C(\mathscr{C}_t \in \mathcal{U} \mid T_{\mathcal{U}} \leq t) \geq q \prod_{s=T_{\mathcal{U}}+1}^{t} \mathbb{P}_C(\mathscr{C}_s = \mathscr{C}_{s-1} \mid \mathscr{C}_{T_{\mathcal{U}}} = \cdots \mathscr{C}_{s-1}, T_{\mathcal{U}} \leq t),$$

对每个 $C \subset_n \mathbb{Z}^d$ 都成立。根据引理 2.1，存在 $r > 0$，使得对满足 (70) 的每个 U，都有 $\mathbb{P}_U(\mathscr{C}_1 = \mathscr{C}_0) \geq r$。因此，根据应用于 $s \in [\![T_{\mathcal{U}} + 1,\ t]\!]$ 的强马尔可夫性质，以及已知 $\{\mathscr{C}_{T_{\mathcal{U}}} = \cdots = \mathscr{C}_{s-1}, T_{\mathcal{U}} \leq t\}$ 发生时 $\mathscr{C}_{s-1}$ 满足 (70)，我们有

$$\mathbb{P}_C(\mathscr{C}_s = \mathscr{C}_{s-1} \mid \mathscr{C}_{T_{\mathcal{U}}} = \cdots \mathscr{C}_{s-1}, T_{\mathcal{U}} \leq t) \geq r,$$

这意味着所声称的界成立，且带有 $p = qr$。 □

6.2. **引理3.2的应用**。 引理3.2指出，如果 $\mathscr{C}_0$ 中的某个元素包含一个属于含有 m 个或更少元素的连通分量的元素，那么 $\mathscr{C}_1$ 通常包含更少的这类元素。接下来的结果指出，因此，$\mathscr{C}_n$ 通常不含这类元素。换言之，事件

$$\mathsf{NoSmallComp}_t := \cap_{x \in \mathscr{C}_t}\{|\mathrm{comp}_{\mathscr{C}_t}(x)| > m\}$$

对于 $t = n$通常会发生。（我们继续用$\mathrm{comp}_C(y)$表示在 $C \subset \mathbb{Z}^d$中与 $y \in \mathbb{Z}^d$ 连通的元素集合。）

命题 6.1. 存在 $p > 0$ 使得

$$\mathbb{P}_C(\mathsf{NoSmallComp}_n) \geq p, \tag{71}$$

对每个 $C \subset_n \mathbb{Z}^d$。

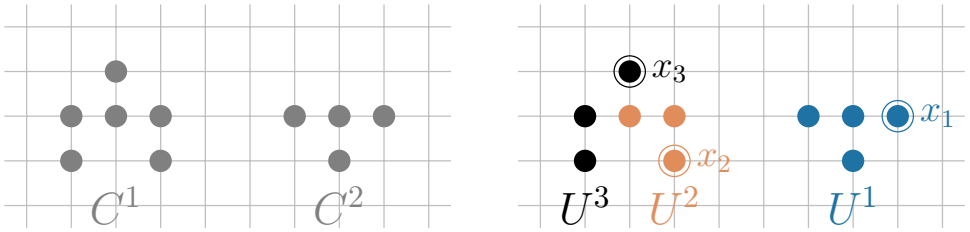

图6. 引理 6.2 的一个示例，其中 $(d,\ m,\ n) = (2,\ 2,\ 10)$。当一个集合 C（左侧）的连通分量 (C^1 和 C^2) 各自包含超过 m 个元素时，可以将它们划分为 $(U^j)_j$，使得每个 U^j 包含 $m+1$ 到 $2m+1$ 个元素，并含有一个元素 x_j（圈出），该元素 (i) 与一条不含 U_k 的射线相邻，且 (ii) 对每个 $k \geq j$，其位置至少与 x_k 一样靠右。这些性质使得各分划可以逐一向右移动（图 7）。

证明。定义 τ 为 NoSmallComp$_t$ 首次发生的时刻 $t \geq 0$。回想一下事件“耗尽小连通分量”$_t$，其定义为：$\{|\mathscr{R}_{t+1}| < |\mathscr{R}_t|\}$ 就 $\mathscr{R}_t$而言，属于$\mathscr{C}_t$ 的连通分量且其中元素不超过 m 个的 $x \in \mathscr{C}_t$ 的集合 (27)。关键观察是

$$\mathbb{P}_C(\tau \leq n) \geq \mathbb{P}_C(\cap_{t<\tau} \mathsf{DepleteSmallComps}_t). \tag{72}$$

确实，因为 $|\mathscr{R}_0| \leq n, |\mathscr{R}_t|$ 在时间 τ之前连续最多只能减少 n 次。如果 $\{\tau \leq n\}$ 发生，那么只需在时间 n之前一直待在 $\mathscr{C}_\tau$，就可以实现 NoSmallComp$_n$。

引理3.2意味着存在 $q_1 > 0$，使得

$$\mathbb{P}_C(\cap_{t<\tau} \mathsf{DepleteSmallComps}_t) \geq q_1,$$

对每个 $C \subset_n \mathbb{Z}^d$均成立，而引理 6.1意味着存在 $q_2 > 0$，使得

$$\mathbb{P}_C(\mathsf{NoSmallComp}_n \mid \tau \leq n) \geq q_2.$$

事实上，引理 6.1适用于由 $\mathbb{Z}^d$的 n元子集构成的集合 $\mathcal{U}$，其中的元素仅属于含有超过 m 个元素的连通分量。显然，该集合满足（70）。将这些界与（72）结合，即可得到当 $p = q_1 q_2$时的（71）。□

6.3. **不存在小连通分量的集合的一个性质**。仅由含有超过 m 个元素的连通分量组成的集合，其优点在于，这些连通分量可以划分为若干分划，使其能够组成直线，然后逐个“放上跑步机”（图6和图7）。

引理 6.2。如果构型 $C \subset_n \mathbb{Z}^d$ 中的每个元素都属于一个含有超过m个元素的连通分量，那么 C存在一个划分为$(U^j)_{j=1}^J$的划分，其中每个分划都满足 $|U^j| \in [\![m+1,\ 2m+1]\!]$和直径$(U^j) \leq n$。此外，每个 U^j都包含某个元素x_j，使得对于每个 $i < j$，射线$(x_j + L_{1,\infty})$上都不含 $\cup_{k \geq j} U^k$ 和 $x_j \cdot e_1 \leq x_i \cdot e_1$。

证明。 将 C 的连通分量记为 $(C^i)_{i \in I}$。注意，由于 $|C^i| \geq m+1$，存在 $|C^i|$ 的一个划分 $(n_l^i)_l$，使得 $n_l^i \in [\![m+1, 2m+1]\!]$。构造 U^j 时，若 $x_j \in C^{i_j} \setminus \{U^k\}_{k<j}$ 在 $C \setminus \{U^k\}_{k<j}$的元素中具有最大的$e_1$ 分量——且我们恰好在之前的 l_j 轮中从分量 i_j 分配了元素——那么我们就将 x_j 以及 $C^{i_j} \setminus \{U^k\}_{k<j}$ 中任何其他的 $n_{l_j+1}^{i_j} - 1$ 个元素分配给 U^j。根据构造，$(U^j)_j$ 具有所述性质。□

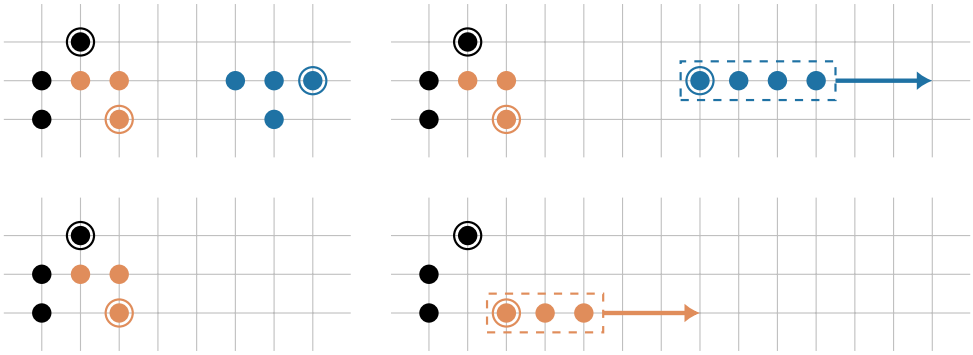

图 7。图6的延续。如果一个集合没有小的连通分量，那么可以将这些连通分量划分为若干部分，并利用引理 6.4 将这些部分逐一向右进行跑步机传输。

6.4. **踏车引理**。 下一个结果表明，若 C 包含一条由 $m+1$ 个或更多元素组成的线段，且其旁边有一条未占用的线段，那么前一条线段通常会平移到后一条线段上。我们将此称为踏车，因为可以先在该线段的“尾部”进行激活，再传输至其“头部”来实现。回忆一下，对于整数 $i \le j$、$L_{i,j}$ 表示线段$\{ke_1 : \ i \le k \le j\}$。例如，若 $x \in \mathbb{Z}^d$，则 $(x+L_{0,k-1})$ 是长度为 k 的线段，其端点位于 x 和 $x+(k-1)e_1$。

引理 6.3 (第一踏车引理)。 假设 $C \subset_n \mathbb{Z}^d$ 包含$(x+L_{0,\ k-1})$，其中 $x \in \mathbb{Z}^d$ 和$k \ge m+1$。存在 $p>0$，使得如果对于某个 $l \in [\![1、\ m]\!]$，线段$(x+L_{k,\ k+l-1})$ 与C不相交，那么

$$\mathbb{P}_C\left(\mathscr{C}_1 = \left(C \setminus (x+L_{0,k-1})\right) \cup (x+L_{l,k+l-1})\right) \ge p. \tag{73}$$

证明。 要将 $L_{0,\ k-1}$ 沿 e_1 方向的若干个位置进行跑步机传输，我们激活“尾部”$(x+L_{0,\ m-1})$，并将 l 个元素传输到“头部”$(x+L_{l,\ k+l-1})$。我们将剩余的 $m-l$ 个元素送回$(x+L_{l,\ m-1})$。因此，所讨论的概率至少为

$$\mathbb{P}_C\left(\mathcal{A}_0 = (x+L_{0,m-1}), \mathcal{T}_0 = (x+L_{l,m-1}) \cup (x+L_{k,k+l-1})\right) > 0.$$

事实上，由于 $(x+L_{0,\ k+l-1})$ 的直径至多为 $2n$，引理 2.1 给出了一个下界 $p>0$，且仅取决于 n。□

第二踏车引理将 $m+1$ 或更多个 C 元素组成一个线段，然后反复应用引理 6.3，使该线段沿跑步机移动所需的距离（图 7）。

引理 6.4 （第二踏车引理）。设 $r \in \mathbb{Z}_{\ge 2m}$ 和 $C \subset_n \mathbb{Z}^d$。存在 $p>0$，使得若 $U \subseteq C$ 的直径至多为 n，满足 $|U| \in [\![m+1、\ 2m+1]\!]$，并且包含一个 x，使得 $(x+L_{1,\ r})$ 与 C 不相交，则

$$\mathbb{P}_C(\mathscr{C}_{r+2} = (C \setminus U) \cup (x+L_{r-|U|+1,r})) \ge p^r. \tag{74}$$

证明。首先，我们用两步将 U 组成线段 $(x+L_{0,\ |U|-1})$。具体而言，我们断言存在 $q_1>0$，使得

$$\mathbb{P}_C\left(\mathscr{C}_2 = (C \setminus U) \cup (x+L_{0,|U|-1})\right) \ge q_1. \tag{75}$$

事实上，我们在第一步中激活任意 m 个 $U \setminus \{x\}$ 元素，并将它们传输到$(x+L_{1,m})$。在第二步中，我们激活其余的 $|U|-(m+1)$ 个元素以及$(x+L_{m-(|U|-(m+1))+1,m})$，从而总共激活 m 个元素，并将这些元素传输到$(x+L_{m-(|U|-(m+1))+1,|U|-1})$。（注意

如果 $|U| = m+1$，那么第二步会使 $\mathscr{C}_1$ 保持不变。）这表明式（75）中的概率为正。由引理 2.1 可得该界，因为前面的步骤蕴含着对 $j \in \{0,1\}$，$\operatorname{diam}(\mathcal{A}_j \cup \mathcal{T}_j) \le 2n$。

其次，我们用 $s = \lceil \frac{r-|U|+1}{m} \rceil \le r$ 额外步数将线段 $(x + L_{0,\ |U|-1})$ 推进到 $(x \pm L_{r-|U|+1,\ r})$，其中 $\lceil \cdot \rceil$ 表示上取整函数。对 $[\![2,\ s+1]\!]$ 中的每个时刻应用马尔可夫性质，并结合引理 6.3，可得

$$\mathbb{P}_C\big(\mathscr{C}_{s+2} = (C \setminus U) \cup (x + L_{r-|U|+1,r}) \,\big|\, \mathscr{C}_2 = D\big) \ge q_2^s, \tag{76}$$

其中，D 表示（75）中的构型，且 q_2 为正。

最后，我们注意到，若 $\mathcal{U} = \{(C \setminus U) \cup (x + L_{r-|U|+1,\ r})\}$，则由（76）可知，$\{T_{\mathcal{U}} \le r+2\}$ 发生的概率至少为 q_2^s。容易验证，$\mathcal{U}$ 满足停留引理（引理 6.1）的假设，因此可得

$$\mathbb{P}_C(\mathscr{C}_{r+2} = (C \setminus U) \cup (x + L_{r-|U|+1,r}) \mid T_{\mathcal{U}} \le r+2) \ge q_3^r,$$

对于 $q_3 > 0$，对每个 $C \subset_n \mathbb{Z}^d$。结合前面的三个界可得

$$\mathbb{P}_C\big(\mathscr{C}_{r+2} = (C \setminus U) \cup (x + L_{r-|U|+1,r}\big) \ge q_1 q_2^s q_3^r.$$

因此，(74)在 $p = q_1 q_2 q_3$ 下成立。 □

6.5. **定理 1.3 的证明**。 按照图 7，我们将使用引理 6.1 把 $\mathscr{C}_0$ 分成含有超过 m 个元素的连通分量，然后沿 e_1 方向反复对每个分量的各个分划部分进行跑步机操作，直至它们满足 (7)。我们将按某种顺序对这些部分进行跑步机操作，以确保每个部分都有一条“开放通道”。

定理 *1.3* 的证明。 给定 $\mathscr{C}_n$，当 $\mathsf{NoSmallComp}_n$ 发生时，$\mathscr{C}_n$ 可以划分为 $(U^j)_{j=1}^J$，且该划分具有引理 6.2 所述的性质。基于此划分，我们定义事件

$$\mathsf{Tread}_j = \left\{\mathscr{C}_{s_j} = (\mathscr{C}_{s_{j-1}} \setminus U^j) \cup V^j\right\},$$

其中

- $r_j = (r+n)(n-j)$ 表示我们对第 j 个部分进行跑步机操作的距离；
- $s_j = \sum_{i \le j}(r_i + 2)$ 是引理 6.4 按相应方式对前 j 个分划进行跑步机处理所需的时间；并且
- $V^j = (x_j + L_{r_j - |U^j|+1,\ r_j})$ 是对第 j 个分划进行跑步机处理后得到的线段。

我们声称

$$\begin{aligned}\mathbb{P}_C\left(\mathscr{C}_{2n^2r} \in \mathsf{CanPersist}_r\right) &\ge \mathbb{E}_C\left[\mathbb{P}_{\mathscr{C}_n}\left(\mathscr{C}_{2n^2r-n} \in \mathsf{CanPersist}_r\right)\mathbf{1}_{\mathsf{NoSmallComp}_n}\right] \\ &\ge \mathbb{E}_C\left[\mathbb{P}_{\mathscr{C}_n}\left(\cap_{j\le J}\mathsf{Tread}_j \cap \mathsf{Stay}\right)\mathbf{1}_{\mathsf{NoSmallComp}_n}\right],\end{aligned} \tag{77}$$

其中 Stay 是 $\cap_{t=s_J+1}^{2n^2r-n}\{\mathscr{C}_t = \mathscr{C}_{t-1}\}$ 的缩写。[7]第一个不等式由塔式法则以及将马尔可夫性质应用于时刻 n 得出。第二个不等式成立，因为当 $\cap_{j\le J}\mathsf{Tread}^j$ 和 Stay 发生时，我们有

$$\mathscr{C}_{2n^2r-n} = \mathscr{C}_{s_J} = \cup_{j\le J} V^j.$$

当该事件发生时，$\{\mathscr{C}_{2n^2r-n} \in \mathsf{CanPersist}_r\}$ 也会发生，因为 $(V^j)_{j\le J}$ 满足 (7)。事实上，引理 6.2 中 $(U^j)_{j\le J}$ 的性质包括 $|U^j| \in [\![m+1,\ 2m+1]\!]$，因此 $|V^j|$ 也具有同样的性质。此外，对于 $k < l$，$x_l \cdot e_1 \le x_k \cdot e_1$，这意味着

$$\operatorname{dist}(V^j, V^{\neq j}) \ge \min_{i\neq j} |r_i - |U_i| - r_j| \ge r.$$

[7]简单计算可得，$r \ge 2n$ 足以保证 $s_J + 1 \le 2n^2r - n$，因为 $J \le n$。

我们对（77 ），其中包含三个引理；我们将在最后验证它们的假设。首先，引理 6.1 指出 t

$$\mathbb{P}_C(\mathsf{NoSmallComp}_n) \geq q_1.$$

其次，根据在时刻s_{j-1} 应用的马尔可夫性质和引理 6.4，我们有

$$\mathbb{P}_{\mathscr{C}_n}\left(\mathsf{Tread}_j \,\middle|\, \cap_{i<j}\mathsf{Tread}_i\right) \geq q_2^{r_j}\mathbf{1}_{\mathsf{NoSmallComp}_n}. \tag{78}$$

第三，根据引理 6.1 和平凡界 $2n^2r - n - s_J \leq 2n^2r$，我们有

$$\mathbb{P}_{\mathscr{C}_n}\left(\mathsf{Stay} \,\middle|\, \cap_{j\leq J}\mathsf{Tread}_j\right) \geq q_3^{2n^2r}\mathbf{1}_{\mathsf{NoSmallComp}_n}. \tag{79}$$

（q_1、q_2、q_3 均为正数。）

这些界合起来意味着

$$\mathbb{E}_C\left[\mathbb{P}_{\mathscr{C}_n}\left(\cap_{j=1}^{J}\mathsf{Tread}_j \cap \mathsf{Stay}\right)\mathbf{1}_{\mathsf{NoSmallComp}_n}\right] \geq q_1 q_2^{\sum_j r_j} q_3^{2n^2r}, \tag{80}$$

其中， q_2 上的指数满足

$$\sum_{j\leq J} r_j \leq n^2(r+n) \leq 2n^2r.$$

第一个不等式利用了这样一个事实： $J \leq n$ 和 $r_j \leq (r+n)n$；第二个不等式利用了假设 $r \geq 2n$。将此界代入 (80)，再将 (80) 代入 (77)，可得

$$\mathbb{P}_C\left(\mathscr{C}_{2n^2r} \in \mathsf{CanPersist}_r\right) \geq q_1(q_2q_3)^{2n^2r}.$$

特别地，(8) 在 $p = (q_1q_2q_3)^{2n^2}$的条件下成立。

仍需 v 证我们在推导 (78) 和 (79 时所用的引理 6.4 和引理 6.1 的假设).

- 引理 6.4 的假设中，只有一项并非显然满足——即 U^j 包含一个 x，使得 $(x + L_{1,\,r_j})$ 与 $\mathscr{C}_{s_{j-1}}$ 不相交。事实上，x_j 满足这一要求，因为根据引理 6.2，射线 $x_j + L_{1,\,\infty}$ 上没有 $\{U^i\}_{i\geq j}$，并且当 $\cap_{i<j}\mathsf{Tread}_i$ 发生时， $\mathscr{C}_{s_{j-1}}$的其余部分由 $\{V^i\}_{i<j}$构成，后者与 x_j至少相距 $r_{j-1} - n > r_j$。

- 我们使用引理 6.1 时，需要对每个 $t \in [\![s_J + 1, 2nr^2 - n]\!]$， $\mathscr{C}_t$ 都包含一个满足 (10) 的 m 元子集。但当 $\cap_{j\leq J}\mathsf{Tread}^j$ 和 $\cap_{u=s_J+1}^{t-1}\{\mathscr{C}_u = \mathscr{C}_{u-1}\}$ 出现时，每个 V^j 都包含这样一个子集。

□

7. 不可约性

本节通过证明 CAT 在 $\mathrm{Prog}_{m,n}$ 上不可约（命题 1.1），完成定理 1.5 的证明。将 CAT 首次到达 $U \subset_n \mathbb{Z}^d$ 的时刻记为 $T_U := \inf\{t \geq 0 : \mathscr{C}_t = U\}$，并进一步将 $T_{\widehat{U}} := \inf\{t \geq 0 : \mathscr{C}_t = U\}$ 记为 。要使 CAT 在 $\mathrm{Prog}_{m,n}$ 上不可约，CAT 必须满足

$$\forall U, V \in \mathrm{Prog}_{m,n}, \quad \mathbb{P}_U\left(T_{\widehat{V}} < \infty\right) > 0. \tag{81}$$

换言之，CAT 可以从任意这样的 U到达任意这样的 V。我们通过证明 CAT 可以从任意 U到达一条直线，并且可以从一条直线到达任意 V来证明这一点。我们继续用 $L_{1,\,n} = \{e_1, 2e_1, \ldots, ne_1\}$ 表示。

命题 7.1(集合到直线)。对于每个 $U \in \mathrm{Prog}_{m,\,n}$,

$$\mathbb{P}_U\left(T_{\widehat{L}_{1,n}} < \infty\right) > 0.$$

命题 7.2 (直线到集合)。 对于每个 $V \in \mathrm{Prog}_{m,\,n}$,

$$\mathbb{P}_{L_{1,n}}\left(T_V < \infty\right) > 0.$$

命题 *1.1* 的证明。 不等式(81)由对$T\widehat{L_{1,\ n}}$和$\widehat{\ }$命题 7.1 和命题 7.2 应用强马尔可夫性质得到。因此，CAT 在 Prog$_{m,\ n}$上不可约。□

我们将　　注意力转向证明命题 7.2，因为命题 7.1 可由先前的结果推出。

命题 *7.1* 的证明。 这可直接由命题 5.7 得出。 □

我们将对 m 使用归纳法来证明命题 7.2。为便于证明，我们预先处理两种情况。接下来的结果讨论第一种情况。

命题 7.3。 对于每个 $V \in \mathrm{Prog}_{n-1,n}$,

$$\mathbb{P}_{L_{1,n}}(T_V < \infty) > 0.$$

证明。 如果 $V \in$ Prog$_{n-1,\ n}$，那么 V是连通的。因此，如果$L_{1,\ n}$与 V相交，那么 CAT 可以一步形成 V。否则，我们可以让$L_{1,\ n}$中的元素在跑步机上移动，直到它们与 V相交。□

下一个结果讨论第二种情况。

命题 7.4. 对于任意 $V \in \mathrm{Prog}_{1,\ n}$,

$$\mathbb{P}_{L_{1,n}}\big(T_V < \infty\big) > 0.$$

我们省略一个简单事实的证明：如果 CAT 可以从 A形成 B，那么只要 $\|a\|$ 足够大，CAT 就可以从 $A\cup\{a\}$形成 $B\cup\{a\}$，对于 $a \in \mathbb{Z}^d$也是如此。

引理 7.1。 设 $A, B \in \mathcal{S}$。如果 $a \in \mathbb{Z}^d$ 的 $\|a\|$足够大，那么

$$\mathbb{P}_A(T_B < \infty) > 0 \implies \mathbb{P}_{A\cup\{a\}}(T_{B\cup\{a\}} < \infty) > 0. \tag{82}$$

命题 *7.4* 的证明。 我们对 $n \geq 2$使用归纳法。命题 7.3 讨论了 $n = 2$的基例。因此，我们假设

$$\forall U \in \mathrm{Prog}_{1,n}, \quad \mathbb{P}_{L_{1,n}}(T_U < \infty) > 0, \tag{83}$$

并且我们旨在证明

$$\forall V \in \mathrm{Prog}_{1,n+1}, \quad \mathbb{P}_{L_{1,n+1}}(T_V < \infty) > 0. \tag{84}$$

我们将把 $(n+1)e_1 \in L_{1,\ n+1}$ 隔离在远离原点的位置。随后，利用归纳假设和引理 7.1，我们可以从 $L_{1,\ n}$构造一个集合 W，使其几乎是 V。具体而言，我们将利用归纳假设构造

$$W = (V \setminus \{v, g\}) \cup \{\ell - e_1\},$$

其中， v 是 V的任意一个非孤立元素，而 g 和 ℓ 分别是 V按字典序排列时的最大元素和最小元素。接下来，我们将分两步把 W 变成 V。

我们可以如下构造 V。

(1) 隔离一个元素。从 $L_{1,\ n+1}$出发，我们沿 e_1 方向让这一对 $\{ne_1, (n+1)e_1\}$ 在 e_1 上运行，直到其中一个（设为 a）的范数 $\|a\|$ 足够大，从而保证 (82) 对$A = L_{1,\ n}$和 $B = W$成立。然后，我们将另一个元素送回 ne_1，以构成 $L_{1,\ n} \cup \{a\}$。

(2) 使用归纳假设。根据设计， $W \in$ Prog$_{1,\ n}$ 因为 W 包含 $\{\ell, \ell - e_1\}$。因此，归纳假设表明，我们可以由 L_n构造出 W。根据引理 7.1，我们实际上可以由 L_1构造出$W \cup \{a\}$，$_n \cup \{a\}$。

(3) 返回 孤立元素。从 $W \cup \{a\}$ 出发，我们在 a 处激活，并传输到 $\ell - 2e_1 \in \partial \quad W$.

(4) 对词典序元素运行跑步机。然后我们对从 $\{\ell - e_1, \ell - 2e_1\}$ 到 $\{g, g + e_d\}$ 的元素对运行跑步机；这是可行的，因为 $\ell - 2e_1$ 和 g 之间存在一条不经过 W 的路径。结果是 $(V \setminus \{v\}) \cup \{g + e_d\}$。最后，我们在 $g + e_d$ 处激活，并传输到 v；这是可行的，因为 v 在 V 中不是孤立的。

□

命题 7.2 的证明中的归纳步骤比命题 7.4 的证明中的归纳步骤更容易。原因是，对于某个 $U \in \mathbf{Prog}_{1,n+1}$，不存在满足 $U \setminus \{u\} \in \mathbf{Prog}_{1,n}$ 的非孤立 $u \in U$。然而，对于每个 $V \in \mathbf{Prog}_{m+1,n+1}$，都存在一个非孤立的 $v \in V$，使得 $V \setminus \{v\} \in \mathbf{Prog}_{m,n}$。（请注意，在后一种情况下，$n+1$ 和 $m+1$ 都递减为 n 和 m。）这由 $\mathbf{Prog}_{m+1,n+1}$ 的定义可知；该定义要求它的每个成员 V 都有一个 $(m+1)$ 元素子集 $\{y_1, \ldots, y_{m+1}\}$，并且该子集尤其满足

$$y_{m+1} \in \partial(V \setminus \{y_{m+1}\}) \quad \text{and} \quad V \setminus \{y_{m+1}\} \in \mathrm{Prog}_{m,n}. \tag{85}$$

换言之，y_{m+1} 就是前面提到的 v。

命题 7.2 的证明。我们对 $m \geq 1$使用归纳法。$m = 1$ 的基例由命题 7.4 得出。因此，我们假设对于所有 $n \geq m + 1$,

$$\forall U \in \mathrm{Prog}_{m,n}, \quad \mathbb{P}_{L_{1,n}}(T_U < \infty) > 0. \tag{86}$$

我们旨在证明，对于所有 $n' \geq m + 2$,

$$\forall V \in \mathrm{Prog}_{m+1,n'}, \quad \mathbb{P}_{L_{1,n'}}(T_V < \infty) > 0, \tag{87}$$

事实上，我们只需处理 $n' \geq m + 3$，因为命题 7.3 证明了当 $n' = m + 2$时(87)成立。

假设 $n' \geq m + 3$。我们可以如下构造 V。

(1) 隔离一个元素。从$L_{1,n'}$出发，我们沿 e_1 方向使这对元素 $\{(n'-1)e_1, n'e_1\}$ 像跑步机一样移动，直到其中一个（例如 a处的元素）的半径足够大，从而确保(82)对$A = L_{1,n'}$ 和 $B = V \setminus \{y_{m+1}\}$ 成立。然后将另一个元素送回 $n'e_1$，构成$L_{1,n'-1} \cup \{a\}$。注意，要在不移动其他元素的情况下让这对元素像跑步机一样移动，我们利用了 $n' \geq m + 3$这一事实。具体来说，每一步中，我们激活这对元素中的一个，并 m 个其他元素。由于至少有 $m + 1$ 个元素不属于这对元素，我们可以将所激活的 m 个元素传输回它们被激活的位置。(2) 使用归纳假设。由(85), $V \setminus \{y_{m+1}\} \in \mathbf{Prog}_{m,n'-1}$。因此，我们可以将 $n = n' - 1$ 和 $U = V \setminus \{y_{m+1}\}$ 代入归纳假设(86)，从$L_{1,n'-1}$构造 U。事实上，由引理 7.1，我们可以从$L_{1,n'-1} \cup \{a\}$构造 $U \cup \{a\}$。(3) 送回孤立元素。最后，在 $\{y_1, \ldots, y_{m-1}\}$ 和 w处激活，并传输至 $\{y_1, \ldots, y_m\}$，以构成 V。

□

参考文献

[AACD+17] Marta Andrés Arroyo、Sarah Cannon、Joshua J. Daymude、Dana Randall 和 Andréa W. Richa。可编程物质中捷径桥接的随机方法。载于 Robert Brijder 和 Lulu Qian（编），*DNA*计算与分子编程，第122–138页，沙姆，2017年。施普林格国际出版。[BF64] David Blackwell 和 David Freedman。马尔可夫链的尾 σ-域与 Orey 定理。《数学统计年刊》，35(3):1291–1295，1964年。

[BR20] Renaud Bastien 和 Pawel Romanczuk。一个纯粹基于视觉的集体行为模型。《科学进展》，6(6):eaay0792，2020年。[BSR01] Madeleine Beekman、David J. T. Sumpter 和 Francis L. W. Ratnieks。法老蚁觅食行为从无序到有序的相变。《美国国家科学院院刊》，98(17):9703–9706，2001年。[Cal21] Jacob Calvert。调和激活与输运中相变的存在性。arXiv 预印本 arXiv:2110 13893，2021年。。[CDG+19] Sarah Cannon、Joshua J. Daymude、Cem Gökmen、Dana Randall 和 Andréa W. Richa。异质自组织粒子系统中用于分离的局部随机算法。载于 Dimitris Achlioptas 和 László A. Végh（编），《近似、随机化与组合优化：算法与技术》，APPROX/RANDOM 2019，2019年9月20日至22日，美国马萨诸塞州剑桥市麻省理工学院，LIPIcs第145卷，第54:1–54:22页。Schloss Dagstuhl - Leibniz 信息学中心，2019 年。[CDRR16] Sarah Cannon 、 Joshua J. Daymude 、 Dana Randall 和 Andréa W. Richa。自组织粒子系统中用于压缩的马尔可夫链算法。载于《2016年ACM分布式计算原理研讨会论文集》，PODC ’16，第279–288页，美国纽约州纽约市，2016年。计算机协会。[CF14] Rebecca M. Clark 和 Jennifer H. Fewell。沙漠切叶蚁 Acromyrmex versicolor 的蚁群增长从不稳定向稳定转变。《行为生态学与社会生物学》，68(1):163–171，2014年。[CGH21] Jacob Calvert、Shirshendu Ganguly 和 Alan Hammond。调和激活与输运中的坍缩与扩散。arXiv 预印本 arXiv:2110.13895，2021年。[DDG+14] Zahra Derakhshandeh、Shlomi Dolev、Robert Gmyr、Andréa W. Richa、Christian Scheideler 和 Thim Strothmann。变形虫机器人——一种可编程物质新模型。载于《第26届ACM算法与体系结构并行性研讨会论文集》，SPAA ’14，第220–222页，美国纽约州纽约市，2014年。计算机协会。[DVM04] Ruth Daniels、Jos Vanderleyden 和 Jan Michiels。细菌的群体感应与群体迁移。《FEMS微生物学评论》，28(3):261–289，2004年6月。[FMM95] G. Fayolle、V. A. Malyshev 和 M. V. Menshikov。《可数马尔可夫链的构造理论专题》。剑桥大学出版社，1995年。[GG17] Dan Gorbonos 和 Nir S. Gov。利用自适应长程相互作用实现稳定群集。《物理评论E》，95:042405，2017年4月。[GIP+16] Dan Gorbonos、Reuven Ianconescu、James G Puckett、Rui Ni、Nicholas T Ouellette 和 Nir S Gov。昆虫群中的长程声学相互作用：一种自适应引力模型。《新物理学杂志》，18(7):073042，2016年7月。[GLPP17] Shirshendu Ganguly、Lionel Levine、Yuval Peres 和 James Propp。通过竞争侵蚀形成界面。《概率论及其相关领域》，168(1):455–509，2017年。[GLS19] Shirshendu Ganguly、Lionel Levine 和 Sourav Sarkar。通过竞争侵蚀形成大尺度随机结构。《概率年刊》，47(6):3649–3704，2019年。[GP18] Shirshendu Ganguly 和 Yuval Peres。竞争侵蚀具有共形不变性。《数学物理通讯》，362(2):455–511，2018年。[GPvdV+20] Dan Gorbonos、James G. Puckett、Kasper van der Vaart、Michael Sinhuber、Nicholas T. Ouellette 和 Nir S. Gov。自适应长程相互作用驱动昆虫群中的配对形成。《英国皇家学会界面期刊》，17(171):20200367，2020年。[HAC+21] Roy Harpaz、Ariel C. Aspiras、Sydney Chambule、Sierra Tseng、Marie-Abèle Bind、Florian Engert、Mark C. Fishman 和 Armin Bahl。斑马鱼的集体行为源于受基因控制的简单行为模式。《科学进展》，7(41):eabi7460，2021年。[JMAK+20] Jitesh Jhawar、Richard G. Morris、U. R. Amith-Kumar、M. Danny Raj、Tim Rogers、Harikrishnan Rajendran 和 Vishwesha Guttal。噪声诱导的鱼群聚集。《自然·物理》，16(4):488–493，2020年。[KHKH22] Hungtang Ko、Mathias Hadgu、Keyana Komilian 和 David L. Hu。小型火蚁筏不稳定。《物理评论·流体》，7:090501，2022年9月。[LDC+21] Shengkai Li、Bahnisikha Dutta、Sarah Cannon、Joshua J. Daymude、Ram Avinery、Enes Aydin、Andréa W. Richa、Daniel I. Goldman 和 Dana Randall。利用机械诱导相变对主动内聚颗粒物质进行编程。《科学进展》，7(17):eabe8494，2021年。[LL10] Gregory F. Lawler 和 Vlada Limic。《随机游走：现代导论》。剑桥高等数学研究丛书。剑桥大学出版社，2010年。[LWBB19] François A. Lavergne、Hugo Wendehenne、Tobias Bäuerle 和 Clemens Bechinger。具有依赖视觉感知的运动性的活性粒子的群体形成与内聚性。《科学》，364(6435):70–74，2019年。[MBD+19] Daniel R. Moberg、Daniel Becker、Christoph W. Dierking、Florian Zurheide、Bernhard Bandow、Udo Buck、Arpa Hudait、Valeria Molinero、Francesco Paesani 和 Thomas Zeuch。冰 I 的终结。《美国国家科学院院刊》，116(49):24413–24419，2019年。[OG21] Nicholas T. Ouellette 和 Deborah M. Gordon。生物系统集体行为建模的目标与局限。《物理学前沿》，9:687823，2021年。

[PMRT14] Daniel J. G. Pearce、Adam M. Miller、George Rowlands 和 Matthew S. Turner。投影在控制鸟群中的作用。《美国国家科学院院刊》，111(29):10422–10426，2014。[PO14] James G. Puckett 和 Nicholas T. Ouellette。确定昆虫群渐近大种群规模。《英国皇家学会界面期刊》，11(99):20140710，2014。[Pra05] Stephen C. Pratt。蚂蚁 Temnothorax albipennis 通过相遇率进行群体感应。行为生态学，16(2):488–496，2005年1月。[SCD+18] William Savoie、Sarah Cannon、Joshua J. Daymude、Ross Warkentin、Shengkai Li、Andréa W. Richa、Dana Randall 和 Daniel I. Goldman。趋光性超微粒子。人工生命与机器人学，23(4):459–468，2018年。