arXiv:2304.06170v1 [cs.DS] 12 Apr 2023

借助全局联系实现局部性：2-核对模型设定错误的稳定性

Christian Borgs*† 耿昭 *

2023年4月14日

摘要

对于许多随机图模型，相关生灭过程的分析为估计巨型连通分量、k-核的规模，以及流行病暴发的规模和概率等提供了局部采样算法。本文针对 2-核这一特殊情形，研究这些算法何时能够抵御图模型设定错误。我们证明，对于平均度有界且局部收敛的图，在一种较弱的扩张概念下，局部采样算法能够稳健地估计 2-核及其最大连通分量的规模。我们提出的弱扩张概念推广了经典的扩张定义，同时适用于许多研究充分的随机图模型。

我们的方法包含一个两步撒点论证。第一步，我们通过喷洒证明巨型分量内部存在一个非空的 2-核；第二步，我们将这个非空的 2-核作为种子，再进行一次撒点论证，以证明巨型分量包含一个规模与图的大小成线性关系的 2-核。第二步基于一种针对树部分中顶点的新颖着色方案。我们的算法结果源于撒点论证过程中确立的 2-核结构性质。

我们的局部算法的运行时间是常数，与图的大小无关；该常数的取值取决于所需的渐近精度 ϵ。然而，由于局部极限具有存在性，我们的论证无法给出该常数关于 ϵ 的函数依赖关系的任何界限，也无法界定图需要多大才能使渐近加性误差界 ϵ 成立。

*加利福尼亚州伯克利大学电气工程与计算机科学系。电子邮件：{borgs,gengzhao}@伯克利.edu。

†加利福尼亚大学伯克利分校巴卡尔地球数字材料研究所（BIDMaP）。

1 引言

现实世界中的网络往往规模庞大。例如，截至 2022 [29], ，Facebook 网络拥有超过 20 亿用户，而互联网的节点数轻松超过 100 亿 [18]。这些现实世界网络的庞大规模，使分析人员难以将整个图载入内存，甚至难以完整确定整个网络（例如，为研究流行病而对整个人群进行接触者追踪是不可能的）。因此，检验全局性质或估计全局量往往很困难。此类全局量包括最大连通分量的大小、k-核的大小，或给定流行病传播模型下流行病暴发的概率和规模。

解决这一挑战的一种可行方案是开发局部算法，即只依赖有界邻域的算法。局部算法的一般框架是：从图 G 中抽取一个顶点 v，探索其半径为 r 的邻域（记作 $\mathcal{N}_r(v;G)$），然后计算所需量的局部估计。重复执行这一子程序 T 次，并取平均值作为估计。局部算法的优点包括减少内存使用、可并行化以适用于分布式计算场景（每台机器在独立的邻域上进行计算），以及可扩展性（运行时间不随网络规模增长）。

例如，一种用于确定图 G 中最大连通分量（通常称为“巨型分量”）相对大小的自然局部算法如下：反复随机抽取顶点 v，并输出其中属于规模超过某个阈值的连通分量的顶点所占比例（例如，可通过广度优先搜索进行检查）。显然，这种算法并不总能给出近似正确的答案：它无法区分一个大型连通图 G_0 和由两个 G_0 副本组成的图。然而，当 G 来自某种随机图模型（例如，Erd˝os-R´enyi 随机图、非齐次随机图、配置模型等）时，众所周知，在适当选择参数的情况下，该局部算法以高概率给出近似正确的答案。这些模型局部上具有树状结构，可以用分枝过程来近似。

本文分析针对 2-核的局部算法。暂且稍作推广，图 G 的 k-核 $C^{(k)}(G)$ 是 G 的最大子图，其中所有顶点的度都至少为 k。对于许多经典随机图模型，例如 Erd˝os-R´enyi 模型 [34], 、配置模型 [19], 和非齐次随机图 [37], ，可以通过分枝过程近似来分析 2-核（更一般地说，k-核），这启发了一种自然的局部算法：探索随机顶点的邻域。更准确地说，令 $\mathcal{A}^{(k)}(\mathcal{N}_r(v;G))$ 表示事件：$\mathcal{N}_r(v;G))$ 中包含一棵深度为 r 、以 v 为根的 k-正则树。局部算法从输入图中抽取 T 个顶点，然后将事件 $\mathcal{A}^{(k)}(\mathcal{N}_r(v;G))$ 发生的频率作为 k-核相对大小的估计值。稍加推导即可由 [37] 的结果推出：如果输入图来自非齐次随机图模型（又称图子模型），那么这一局部算法确实能很好地近似 k-核的相对大小。

请注意，除了利用这些图在局部上类似树这一性质外， [37] 的分析还高度依赖非齐次随机图模型中的条件独立性——一旦特征选定，各边便相互独立。但现实世界中的网络既不是局部树状的，也不具备非齐次随机图的独立性特征。这就引出了一个问题：

这些局部算法在模型设定错误时有多稳健?

为回答这一问题，我们希望为不断增长的图序列 G_n 找到合适且相当一般的条件，使得这些算法以高概率给出近似正确的答案。具体而言，该条件应适用于上文提到的随机图模型。此外，理想的条件还应推广到非树状模型（例如家庭模型 [42, 8, 45]），必要时可通过修改算法，使其能够处理大量的环。

一个自然的条件是，图实例在局部应当“看起来像”这些模型。在不指定具体模型的情况下，这一想法自然地由如今已成为标准概念的局部收敛 [11, 2], 来刻画；该概念要求序列中有限邻域的分布收敛。

但显然，这样的条件还不够。事实上，在一个有 n 个顶点的 d-正则图上运行任意局部算法，与在 $\sqrt{n}$ 个互不相交的 d 阶、规模为 $\sqrt{n}$ 的图副本上运行该算法，显然会得到本质上相同的结果；然而，第一个图含有一个规模与顶点数成线性关系的连通 k-核（至少在 d 足够大时如此），第二个图却没有。要解决这个问题，我们需要加入一个条件，要求图具有足够的“连通性”，以免局部算法受到误导。一个自然的条件是扩张性，但这似乎要求过高，因为即使只考虑巨型分量，上述模型也都不是扩张图。

这引出了本文的贡献：我们分析了 2-核这一特殊情形。首先，我们提出了上述 2-核局部算法的一种自然推广，使其适用于含有许多环的图；其次，我们证明该算法对模型设定错误具有鲁棒性。更准确地说，只要图序列 G_n 满足两个条件，该算法就是渐近正确的近似算法（见下文定理 1.3）：

1. 该序列是局部收敛的；
2. 它是一个有界平均度的弱扩张图序列。

我们采用标准的局部收敛概念（将在陈述主要定理时回顾），而弱扩张概念则是 [3] 中提出的大集合扩张概念的弱化。在给出其精确定义之前，我们先描述算法。

作为动机，我们注意到，如果 v 位于 r-邻域 $\mathcal{N}_r(v;G)$ 的 2-核中，那么显然它也位于 G 的 2-核中。但即使它在局部并不位于 2-核中，仍有可能位于 G 的 2-核中。事实上，对于局部类似树的图，我们几乎永远无法在局部邻域内确认某个顶点属于 2-核。我们用以下记号来描述这一点：给定 $\mathcal{N}_r(v;G)$，如果存在一个图 G'，使得 $\mathcal{N}_r(v;G)=\mathcal{N}_r(v;G')$，且 v 位于 G' 的 2-核中，那么我们称 v 可能属于 2-核（注意，在从 v 开始的广度优先搜索中，这很容易检查；参见附录，其中给出了一个关于 $\mathcal{N}_r(v;G)$ 中边数为线性时间复杂度的算法）。利用这一子程序，我们描述以下局部算法，用于估计图中 2-核的大小，以及其最大连通分量的大小，$C^{(2)}_{\max}(G)$。在本文中，我们将图 H 的大小 $|H|$ 定义为其顶点数。

算法 1：2-核大小的局部近似

图 $G=(V,E)$，局部性参数 K，迭代次数 T

输出：$C^{(2)}$（G）和 $C^{(2)}_{\max}$（G）的估计大小

1 针对 $t=1,\dots,T$ **do**

2 　均匀随机采样 $v\in V$

3 　广度优先探索 $\mathcal{N}_R(v;G)$，直到最大半径 R，使得 $|\mathcal{N}_R(v;G)|\le K$

4 　如果给定邻域 $\mathcal{N}_R(v;G)$，$I^{(2)}_t=1$ 可能属于 2-核，则设为 $I^{(2)}_t=1$，否则设为 0。

5 　设定 $I^{(2,\infty)}_t=1$ 若 $I^{(2)}_t=1$ 和 $\mathcal{N}_R(v;G)$包含一个与 v的距离为 R 的顶点 u，则为；否则为 0

6 返回 $I^{(2)}:=\frac{1}{T}\sum_{t=1}^{T} I^{(2)}_t$ 作为估计的分数

大小 $C^{(2)}(G)$，并将 $I^{(2,\infty)}:=\frac{1}{T}\sum_{t=1}^{T} I^{(2,\infty)}_t$ 作为 $C^{(2)}_{\max}(G)$的估计分数大小

备注 1.1。（i）在探索邻域时，可以检查是否符合 2-核候选条件（第 4–5 行）。局部探索的具体（伪代码）实现见附录 A 中的算法 2。（ii）可以通过将 T 个独立采样的顶点分散到不同处理单元上来并行化该算法，

不同的计算节点，每个节点都探索一个顶点的邻域。(iii) 在分布式设置中，每个计算节点在最坏情况下的时间复杂度和空间复杂度为 $O(K^2)$，其中 K 通常取为较大的常数。重要的是，复杂度不会随 $|V|$ 扩展。

提出算法后，我们现在定义弱扩张这一宽松概念。我们采用文献 [13], 中对图 $G=(V,E)$ 的 (ϵ,δ)-割的定义：将 V 划分为两个集合 A 和 A^c，使得二者的大小都至少为 $\epsilon|V|$，而 A 与 A^c 之间的边数至多为 $\delta|V|$。

定义 1.2（弱扩张图序列）。若一个可能是随机的图序列 $\{G_n\}_{n\in\mathbb{N}}$ 满足：对于任意 $\epsilon\in(0,1/2)$，都存在 $\delta>0$（取决于 ϵ），使得以高概率（whp） G_n 不存在 (ϵ,δ)-割，则称其为弱扩张图序列。若存在 $d<\infty$ 使得以高概率， G_n 的平均度有界于 d，则称该序列具有有界平均度。

如上所述，这种弱扩张概念是对文献 [3], 提出的大集合扩张的一种小幅推广，唯一的区别在于：在我们的定义中， δ 可以取决于 ϵ，而在文献 [3] 中则不能。事实证明，这一细微修改对结果的适用性影响很大，因为与大集合扩张条件不同，我们的条件适用于典型随机图模型的巨型分量，尤其适用于非齐次随机图 [12, 引理 11.3] 和配置模型 [13, 命题 4.4]的巨型分量。

1.1 主要结果

我们的主要结果表明，对于图序列 G_n，当该序列是具有有界平均度的弱扩展图的局部收敛序列时，算法 1 能渐近正确地估计 2-核的相对大小。事实证明，我们的结果对更一般的图序列也成立，即通过边渗流从前述图序列得到的图序列。

为陈述这些结果，我们需要一些记号。首先，给定一个局部有限图 G（即一个度有限的、可能是无限的图）以及 G 中的一个顶点 v，我们使用 $C(v;G)$ 表示 G 中 v 的连通分量。接下来，给定一个图 G 和参数 $p\in[0,1]$，我们通过以概率 $1-p$ 独立删除每条边来定义一个渗流图 $G(p)$。

最后，我们回顾一下，如果对于所有 r，随机顶点在 G_n 中的 r-邻域的分布都收敛，则图序列 G_n 被称为局部收敛。形式上，令 $\mathcal{G}^*$ 为所有有根、局部有限图的空间。如果两个有根图同构，我们将在 $\mathcal{G}^*$ 中将它们视为相同，其中同构定义为：存在一个顶点集之间的双射，将根映射到根，将边映射到边，并将非边映射到非边。然后，通过设定$d_{loc}(G,G')=\frac{1}{1+R}$ 在 $\mathcal{G}^*$ 上定义距离 d_{loc}，其中 R 是最大的 R，使得 G 和G' 中的根周围半径为 R 的球同构。该度量将 $\mathcal{G}^*$ 变成一个度量空间。随后，如果存在一个关于 $\mathcal{G}^*$ 的测度 μ，使得对于$\mathcal{G}^*$上的所有连续有界函数 f，经验平均 $\frac{1}{n}\sum_{v_n\in G_n}f(G_n,v_n)$收敛[1]到 $\mathbb{E}_\mu[f]$，则称图序列 G_n 局部收敛。为方便记号起见，我们假设 G_n 恰好有 n 个顶点；本文始终采用这一约定。

给定一个从极限 μ 中抽取的样本 (G,o)，接着我们考虑渗流图 $G(p)$，并确定根所在连通分量为无限的概率 $\zeta(p)$，

$$\zeta(p):=\mathbb{P}_{(G,o)\sim\mu}(|C(o;G(p))|=\infty). \tag{1}$$

我们将 ζ 称为渗流函数。注意到 ζ 是非递减的，我们将渗流阈值 $p_c=p_c(\mu)$ 定义为 $p_c:=\inf\{p\in[0,1]:\zeta(p)>0\}$。正如我们将在下文的定理 2.1 中看到的，对于所有 $p\neq p_c$，其大小

[1]如果序列 G_n 是随机的，那么这个量也是随机的，因此我们要求依概率收敛（我们始终假设 μ 是确定性的，从而右侧不具有随机性）。

渗流序列中最大连通分量的大小 $G_n(p)$ 除以 n，当序列 G_n 局部收敛、具有有界平均度且满足我们的弱扩张图假设时，收敛至 $\zeta(p)$。这尤其意味着，p_c 是 $G_n(p)$ 中巨型连通分量出现的阈值。

做好这些准备后，我们就可以陈述主要结果了。

定理 1.3（用于 2-核的局部算法）。设 $\epsilon>0$ 为给定的误差容限。假设 $\{G_n\}_{n\in\mathbb{N}}$ 是一个局部收敛的弱扩张图序列，且平均度有界，并且 $p\neq p_c$。那么存在常数 K、T，使得算法 1 能以高概率估计渗流图 $G_n(p)$ 的 2-核（整个图和/或其最大连通分量）的相对大小，且加性误差不超过 ϵ。也就是说，输入 $G_n(p)$，并令参数 K、T 足够大时，算法 1 的输出满足

$$\left|\frac{1}{n}|C^{(2)}(G_n(p))|-I^{(2)}\right|\le\epsilon \quad and \quad \left|\frac{1}{n}|C^{(2)}_{\max}(G_n(p))|-I^{(2,\infty)}\right|\le\epsilon, \tag{2}$$

每种情况发生的概率至少为 $1-\epsilon$，其中 n 足够大。此外，所需的迭代次数 T 可由某个普适常数 C 乘以 $C\frac{1}{\epsilon^2}\log\frac{1}{\epsilon}$ 界定。

为了证明该定理，我们将分析 G_n 中随机顶点 v 的局部邻域，并考察在给定邻域 $\mathcal{N}_R(v;G)$ 的情况下，一个可能属于 2-核的随机顶点 v 满足什么条件时才确实属于 2-核。正如我们将看到的，当 $R\to\infty$ 时，对于大多数顶点 v 而言，若图是平均度有界的弱扩张图，渐近地这一结论成立。作为实施这一证明策略的副产品，我们将首先证明以下关于 $C^{(2)}(G(p))$ 和 $C^{(2)}_{\max}(G(p))$ 渐近大小的定理。

为陈述这一点，我们定义 2-核对应的渗流函数 ζ。为此，我们考虑从局部收敛序列 G_n 的极限分布 μ 中抽取的一个无限有根图 (G,o)，并考察渗流图 $G(p)$ 的 2-核 $C^{(2)}(G(p))$。将 $C^{(2)}(G(p))$ 划分为其所有有限连通分量的并集 $C^{(2,<\infty)}(G(p))$ 以及其余部分 $C^{(2,\infty)}(G(p))=C^{(2)}(G(p))\setminus C^{(2,<\infty)}(G(p))$，会比较方便。接下来我们定义三个量：渗流后根属于 2-核中一个无限连通分支的概率，

$$\zeta^{(2,\infty)}(p):=\mathbb{P}_{(G,o)\sim\mu}(o\in C^{(2,\infty)}(G(p))), \tag{3}$$

对应有限连通分量的量 $\zeta^{(2,<\infty)}(p;\mu):=\mathbb{P}_{(G,o)\sim\mu}(o\in C^{(2,<\infty)}(G(p)))$，以及这两者之和，可写作

$$\zeta^{(2)}(p):=\zeta^{(2,<\infty)}(p)+\zeta^{(2,\infty)}(p)=\mathbb{P}_{(G,o)\sim\mu}(o\in C^{(2)}(G(p)). \tag{4}$$

为了直观理解这一量，我们注意到，如果 G_n 局部呈树状，那么 G 几乎必然是一棵树；当且仅当在 $G(p)$ 中存在两条从 o 到 ∞ 的不相交路径时，才有 $o\in C^{(2)}(G(p))$。请注意，这尤其意味着根位于 $G(p)$ 的 2-核中的一个无限连通分支内，即 $\zeta^{(2)}(p)=\zeta^{(2,\infty)}(p)$。但一般而言，$\zeta^{(2,<\infty)}(p)$ 和 $\zeta^{(2,\infty)}(p)$ 都会产生贡献。我们将 $\zeta^{(2,\infty)}(p)$ 称为巨型 2-核[2]的渗流函数。

定理 1.3 本质上可归结为以下关于 2-核相对大小收敛到相应渗流函数的陈述。

[2]研究渗流中巨型（即全局）2-核的出现，往往更有启发性，因为 2-核的小连通分量高度依赖于图的局部结构。例如，考虑如下家庭模型：我们将 Erd˝os-R´enyi 随机图中的每个顶点替换为一个三角形，并将原有的边随机、独立地连接到这三个顶点上。对于任意 $p\in(0,1)$，总有固定比例的家庭以 p 的概率在渗流中得以保留，但只有当渗流超过标准渗流阈值时，才可能存在巨型 2-核。

定理 1.4（2-核的大小）。设 $\{G_n\}_{n\in\mathbb{N}}$ 是一个图序列，其图包含 n 个顶点，并在 $n\to\infty$ 时局部收敛到 μ。假设该图序列是一个平均度有界的弱扩张图序列。那么，对于所有 $p\neq p_c$，

$$\frac{1}{n}|C^{(2)}(G_n(p))|\xrightarrow{p}\zeta^{(2)}(p)\qquad and\qquad \frac{1}{n}|C^{(2)}_{\max}(G_n(p))|\xrightarrow{p}\zeta^{(2,\infty)}(p). \tag{5}$$

证明思路。为了证明定理 1.4，我们采用两步撒点论证。第一步，我们通过喷洒来证明巨型分量内部存在一个非空的 2-核。随后，这个非空的 2-核将作为种子，在第二次撒点论证中生成一个线性规模的 2-核。第二步基于一种针对树部分顶点的新颖着色方案。

未来方向。从算法的角度来看，在我们的弱扩张条件下，考虑还有哪些参数可以通过局部算法得到良好近似将会很有意思，其中 $k\geq 3$ 的 k-核大小是一个自然的候选对象。其次，从更具结构性的角度来看，一个有趣的问题是，在渗流下，弱扩张性质是否以高概率对巨型连通分量保持不变。

1.2 相关工作

k-核（或 k-退化性）的概念最早在 [24, 41] 被提出，用于刻画网络的凝聚性和鲁棒性。k-核的规模已在多种随机图模型中得到研究，包括 Erd˝os-R´enyi 模型 [34], 非齐次随机图 [37], 和配置模型 [19]。这些模型具有局部树状结构；当参数高于某个临界阈值时，可以在一个巨型连通分量中观察到 k-核。对于这些模型，众所周知，在边渗流中巨型连通分量是唯一的，其规模可用适当生灭过程的存活概率表示（例如，[14, 12, 15]）。

关于这些巨型分量结果能否在某些不依赖模型的设定下成立，这个问题直到最近才得到研究 [5, 10, 20, 39]。这些研究采用的主要假设是局部收敛（最初是在不同背景下引入的 [2, 11]），以及某种扩张性质。这些论文中有几篇使用了喷洒论证，尽管这种技术当然早已存在，其渊源可追溯到 Erd˝os 和 R´enyi [14]。值得注意的是，当图并非局部上类似于树时，究竟应该用什么来替代生灭过程的存活概率，事先并不明了；事实上，所有这些关于巨型分量大小的结果都假设局部极限的支撑集由树构成。上述问题由 [3] 和 [44], 独立解答，他们指出，当图并非局部上类似于树时，渗流函数 (1) 应发挥存活概率的作用。沿着 [5, 10, 20, 39], 的研究方向，[3] 的作者随后以更弱的假设取代了先前工作的有界度扩张图假设，即假设图是有界平均度的大集合扩张图；他们还证明，许多更贴近现实的模型都满足这一条件，包括优先连接模型 [35, 9][3]。

然而，在本文发表之前，对于缺乏特定建模假设时 k-核的大小，人们知之甚少。我们的结果推广了 [3], 中的条件，并在这一较弱假设下确定了渗流中巨型分量和 2-核的大小。

最后，作为应对超大规模网络所带来挑战的一种方案，网络上的局部算法已得到广泛研究。最早的研究可追溯至 [7, 25]，当时的研究是从分布式计算的角度展开的。研究者建立了下界，以刻画局部性与近似质量之间的权衡（例如，[25, 30, 22, 21, 17]）。近年来，Rubinfeld 等人开展的工作 [38], 推动局部计算算法社区显著发展。该社区同样关注超大规模图中的局部性，但采用了不同的复杂度度量。

[3]遗憾的是，Erd˝os-R´enyi 随机图的巨型连通分量通常不满足这一条件 .

研究者考虑了各种问题，其中往往涉及超大规模网络上的组合优化。例子包括极大独立集 [40, 16], 、极大匹配 [31, 27], 、顶点覆盖 [32, 31], 、距离估计 [28], 、图伸展子图 [33, 23], 、PageRank [6], 以及聚类检测 [36, 26]。

与本研究最密切相关的是 [4], 的工作，他们研究了在网络上局部估计流行病暴发概率和规模的问题，并针对一种感染持续时间为确定值的流行病模型提出了一种具有理论保证的算法；该算法适用于底层接触网络为有界平均度、局部收敛的大型集合扩张图的情形。从研究思路上看，我们的工作与这项研究密切相关。

2 预备知识

弱扩张与抵御攻击的鲁棒性。我们首先指出，对于任意一列弱集合扩张图，$\frac{1}{n}|C_{\max}(G_n)| \to 1$ 以概率收敛。事实上，固定 $\epsilon \in (0, 1/2)$，并假设 $|C_{\max}(G_n)| \leq (1-\epsilon)n$。根据弱扩张的定义，我们可以找到一个 δ，使得 $C_{\max}(G_n)$ 的边界以高概率至少为 δn。但 $C_{\max}(G_n)$ 的边界为空，矛盾。

同样的论证表明，对于任意容差参数 $\epsilon \in (0, 1/2)$，都存在一个“安全余量” $\delta > 0$，使得以高概率，除非对手至少能删除 δn 条边，否则任何恶意攻击者都无法使巨型分量的规模减少超过 ϵn个顶点（抵御恶意攻击的鲁棒性）。最后不难看出，这两项性质反过来都蕴含弱扩张，这表明弱扩张等价于以下两个陈述：(a) 巨型分量包含除 $o(n)$ 个顶点之外的所有顶点；(b) 巨型分量能够抵御恶意攻击。

注意，这种针对对抗性攻击的鲁棒性意味着其针对随机边删除的鲁棒性，即，它表明，对于任意 $\epsilon \in (0, 1/2)$，存在 $\delta > 0$，使得渗流图 $G_n(p)$ 中巨型分量的大小以高概率大于 $(1-\epsilon)n$，前提是 $p > 1 - \frac{\delta}{2d}$。
渗流中巨型连通分量的大小。对于稍强一些的扩张概念，[3]中证明了以下定理。事实证明，[3] 中的证明可以轻易推广到我们所用的扩张概念，从而得到以下定理。我们在附录 C 中概述其证明。

定理 2.1. 设 $\{G_n\}_{n\in\mathbb{N}}$ 是一个具有有界平均度的弱扩展图局部收敛序列，其极限分布为 $\mu \in \mathcal{G}^*$。那么对于所有 $p \neq p_c$，$\frac{1}{n}|C_{\max}(G_n(p))| \xrightarrow{p} \zeta(p)$。此外，$\zeta$除可能在 p_c 处外均连续。

备注 2.2. 鉴于我们关于弱扩张以及抵御随机删除边的鲁棒性的讨论，容易看出，在定理的假设下，$\zeta(p) \to 1$ 当 $p \to 1$ 时成立，因此特别有$p_c < 1$。

3 2-核相对大小的收敛

现在我们将概述主要命题——命题 3.4——的证明，该命题讨论了 $G_n(p)$ 中 2-核相对大小的收敛。我们将重点关注 2-核的巨型连通分量 $C^{(2)}_{\max}(G_n(p))$；小连通分量（例如孤立环等）是局部的，将在附录 F 中使用标准技术处理。

3.1 在 p_c 以上的巨型分量中出现非空 2-核

我们从以下命题开始，该命题指出，对于 $p > p_c$，以正概率，从极限分布 μ 中抽取的图的根在经过渗流后仍处于无限 2-核中。

命题3.1（临界点以上无限 2-核的出现）。对于 μ，若其为弱扩张图的一个局部极限，则我们有 $\zeta^{(2,\infty)}(p)>0$，当 $p>p_c$；并且 $\zeta^{(2,\infty)}(p)=0$，当 $p<p_c$。

证明概要。为证明该命题的非平凡部分（p_c以上的陈述），我们利用局部极限的性质，即遍历性：根位于 2-核的无限连通分支中的概率非零，当且仅当存在一个与根的距离有限且位于无限 2-核中的顶点的概率非零，见附录G.1 中的推论 G.2。为了将这一概率与局部极限中的无限连通分支的概率联系起来，我们将使用以下恒等式，

$$\mathbb{P}\left(o\in\bigcup_{r\in\mathbb{N}}\mathcal{N}_r(C^{(2,\infty)}(G(p)))\right)=\mathbb{P}(|C(o;G(p))|=\infty), \tag{6}$$

其中，顶点集合 $A\subseteq V$ 的 r-邻域 $\mathcal{N}_r(A;G)$ 就是 $\bigcup_{v\in A}\mathcal{N}_r(v;G)$。该恒等式也在附录 G.1 中得到证明，并且与以下事实有关：对于弱扩展图的局部极限，在 p_c 以上不可能只有一条通向无穷远的路径。□

鉴于我们已经知道，在 p_c 之上存在唯一一个线性大小的巨型连通分量（定理 2.1），定理的第二个结论表明，在 G_n 上方于临界点以上进行渗流，也应留下一个线性大小的巨型 2-核，其渐近大小为 $n\zeta^{(2,\infty)}(p)$。我们将在下一小节证明这一点，而在这里，我们只证明以下命题，其结论较弱：在 p_c 上方，巨型连通分量包含一个非空的 2-核。这个非空的 2-核之后将作为种子，用于另一场喷洒论证，以证明线性大小的 2-核存在。

命题 3.2 (巨型连通分量的非平凡 2-核)。对于 $p>p_c$, $C^{(2)}(C_{\max}(G_n(p)))\neq\emptyset$ 以高概率 .

为概述该命题的证明，我们需要一些记号。回顾一下，$\zeta^{(2,\infty)}$ 是根据 $C^{(2,\infty)}(G)$定义的，后者是图 G的2-核的无限连通分支的并集。令 $\mathcal{N}_\ell^+(v)$表示这样得到的图：从 $\mathcal{N}_\ell(v)$出发，在距离 ℓ from v的每个顶点上植入一条无限射线。然后，我们将 $C^{(2,\ell)}(G)$定义为 $V(G)$中的顶点 v v $v\in$ $C^{(2,\ell)}$ G ℓ 2 的集合，使得 $v\in C^{(2,\infty)}(\mathcal{N}_\ell^+(v))$。换句话说，如果关于-邻域的局部信息无法排除它属于无限-核，则( )。

由于 $C^{(2,\ell)}(G_n(p))$ 中的每个顶点都属于大小至少为 ℓ 的连通分量，而且除最大的连通分量外，所有此类连通分量的总大小在 ℓ 充分大时可以任意小（这是巨型分量唯一性的推论；更准确地说，是因为 $|n^{-1}C_{\max}(G_n(p))|\to_p\zeta(p)$），因此我们可以保证 $|C^{(2,\ell)}(G_n(p))\setminus C_{\max}(G_n(p))|\le\epsilon n$，对于任意 $\epsilon>0$，只要我们选择足够大的 ℓ 即可。$C^{(2,\ell)}(C_{\max}(G_n(p)))$ 大小的以下下界可直接得出；完整论证见附录 G.1。

引理 3.3。对于任意 $\epsilon>0$ 和 $p\in(p_c,1]$，当 ℓ 充分大时，我们有

$$\frac{1}{n}|C^{(2,\ell)}(C_{\max}(G_n(p)))|\ge\zeta^{(2,\infty)}(p)-\epsilon \qquad whp. \tag{7}$$

引理 3.3 表明，巨型分量中的顶点有相当大比例在局部看来属于 2-核的一部分。我们或许可以直接论证，这些顶点中的大多数彼此之间连接错综复杂，因此确实属于 2-核。我们不知道如何直接证明这一点，因此先通过第一次喷洒论证来证明命题 3.2，然后利用巨型分量中的非空 2-核作为种子，进行第二次更复杂的喷洒论证。

命题 3.2 的证明。固定 $p'\in(p_c,p)$。我们可以假设 $C_{\max}(G_n(p'))$ 是一棵树；否则，它的 2-核已经非空。由引理 3.3 可知，对于充分大的 ℓ，我们有

$$\frac{1}{n}|C^{(2,\ell)}(C_{\max}(G_n(p')))|\ge\epsilon:=\zeta^{(2,\infty)}(p')/2>0 \tag{8}$$

以高概率成立，并且我们可以以该事件为条件。考虑一个大小为 $\lfloor \epsilon n/2 \rfloor$ 的连通子图 $H \subseteq C_{\max}(G_n(p'))$，例如从树的一个叶节点出发，通过广度/深度优先搜索构造。将 H （即所有这些顶点及其关联的边）从 $C_{\max}(G_n(p'))$ 中移除后，剩下一个森林，也就是由顶点 $C_{\max}(G_n(p'))\setminus H$ 上若干棵互不相交的树组成的集合。

此时，我们定义以下术语，稍后将会发现它们很有用。考虑一个有限连通图 G 以及一个包含 G 的 2-核（可能为空）的非空连通子图 $H \subseteq G$。那么 $G\setminus H$ 是一个森林，其中每棵树都有一个唯一的顶点，通过该顶点连接到 H 中的 G。我们将这个顶点视为其根，并将树朝向根定向（我们称此方向为下游）。我们将 $u \notin H$ 的下游记为 $\mathsf{ds}(u)$，定义为严格位于 u 与 H之间的顶点集合；而 u 的上游记为$\mathsf{us}(u)$，是满足 $u \in \mathsf{ds}(v)$ 的顶点 $v \notin H$的集合。 $u \notin H$朝向H 在 G中的深度，记为 $\mathsf{depth}_H(u;G)$，递归定义为：如果 u 是 G 中的叶节点（即度为 1），则其值为1；否则为$1+\max_{v\in \mathsf{us}(u)} \mathsf{depth}_H(v;G)$。注意，从 G 中移除 H后会留下一个森林。每棵树都有一个深度 h 最大的唯一顶点，而该顶点恰好是根。图1对此进行了说明。

现在我们回到 $G = C_{\max}(G_n(p'))$ 的情形，其中 $H \subseteq C_{\max}(G_n(p'))$ 的大小为 $\lfloor \epsilon n/2 \rfloor$。令 F 为 $C_{\max}(G_n(p'))\setminus H$ 中由大小至少为 ℓ 的树组成的子森林。那么，任何不在 H或 F中的顶点$v \in C_{\max}(G_n(p'))$，其整个上游都必须包含在 $\mathcal{N}_\ell(v;G_n(p'))$ 的内部；根据 $C^{(2,\ell)}$的定义，这意味着它不可能属于 $C^{(2,\ell)}(C_{\max}(G_n(p')))$。示意图见图1。换句话说， $C^{(2,\ell)}(C_{\max}(G_n(p')))$ 中的所有顶点都必须属于 H 或 F。因此，

$$|V(F)| \geq |V(F) \cap C^{(2,\ell)}(C_{\max}(G_n(p')))| \geq |C^{(2,\ell)}(C_{\max}(G_n(p')))| - |V(H)| \geq \epsilon n - \frac{\epsilon n}{2} = \frac{\epsilon n}{2}. \quad (9)$$

这意味着 H 和 F 都至少包含 $\epsilon n/3$ 个顶点。根据 G_n的弱扩张性质， G_n 中 H 与 F 之间的最小割至少为 δn ，其中 δ 仅取决于 ϵ；由门格尔定理，这转化为 G_n中连接 H与 F的 δn 条不相交路径。由于 $|E(G_n)| \leq dn/2$ 以高概率成立，¯这些路径中至少有 $\delta n/2$ 条长度不超过 $L = d/\delta$。另一方面， F至多由 n/ℓ 棵树组成，因此，在 $G_n(p')$ 中从 F 到 H 的不相交路径数量至多为 n/ℓ （回想一下，假设 $C_{\max}(G_n(p'))$ 是一棵树）。选择 $\ell > 4/\delta$ 使得 $n/\ell < \delta n/4$，从而在 G_n 中至少留下 $\delta n/4$条连接 H与 F的不相交路径，且这些路径并未完全包含在树 $C_{\max}(G_n(p'))$ 中。

现在以概率 $\beta = 1 - \frac{1-p}{1-p'}$撒入边，从而从 $G_n(p')$ 获得 $G_n(p)$；也就是说，获得 $G_n(\beta)$的一个独立副本，并将 $G_n(p)$ 设为包含 $G_n(p')$ 和 $G_n(\beta)$的并集（即其中的边）。在 $G_n(\beta)$ 中，长度至多为 L 的 $\delta n/4$ 条不相交路径均未出现的概率，上界为 $(1-\beta^L)^{\delta n/4} \to 0$，因为 $n \to \infty$。因此，以高概率，在 $C_{\max}(G_n(p))$ 中，某对顶点 $u \in H$ 和 $v \in F$之间存在两条不相同的路径，这意味着 $C_{\max}(G_n(p))$ 不可能是一棵树，并且必须具有非空的 2-核。 □

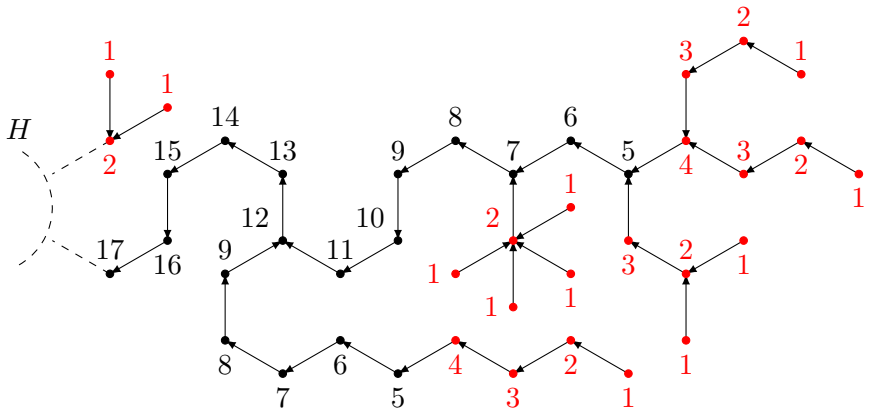


图1：森林 $G\backslash H$ 沿下游方向定向，顶点标注了其朝向 H 的深度。对于 $\ell=4$ 的选择，红色顶点不可能属于 $C^{(2,\ell)}(G)$。

3.2 渗流中 2-核相对大小的收敛

本节概述 $G_n(p)$ 的巨型2-核相对大小概率收敛的证明，重点讨论下界，因为上界可通过使用局部收敛的标准论证得到，并不需要扩张性质。许多细节将在附录中给出，包括巨型 2-核上界的证明（附录E），以及整个 2-核的大小（附录 F）。所需的下界由以下命题给出。

命题 3.4（巨型 2-核大小的下界）。对于任意 $\epsilon>0$ 和 $p\neq p_c$，我们有

$$\frac{1}{n}|C^{(2)}_{\max}(G_n(p))|\geq\frac{1}{n}|C^{(2)}(C_{\max}(G_n(p)))|\geq\zeta^{(2,\infty)}(p)-\epsilon \qquad whp. \tag{10}$$

为证明该命题，我们引入以下着色方案（示意图见附录 G.2 中的图2）。设 $G=(V,E)$ 是一个有限连通图，并设 $H\subseteq G$ 包含 $C^{(2)}(G)$。根据推论 3.2，只需考虑 $C^{(2)}(G)\neq\emptyset$。使用局部化参数 $\ell\in\mathbb{N}$，按如下方式为 G 的顶点和边着色：

1. 若 $\mathrm{depth}_H(v;G)\leq\ell$，则将顶点 $v\notin H$ 染成红色；若 $\mathrm{depth}_H(v;G)>2\ell$ 且 $(\mathrm{depth}_H(v;G)\bmod 2\ell)\in\{1,2,\ldots,\ell\}$，则将其染成紫色；否则保持黑色。2. 将深度连续且颜色相同的顶点之间的每条边染成与端点相同的颜色。这样便在 $G\backslash H$ 中形成红色和紫色的连通分量，我们称之为着色线段。若一个线段包含深度可被 ℓ 整除的顶点，则称其为完整；否则称为不完整。3. 对于每个不完整的紫色线段，将其顶点和边的颜色恢复为黑色。对于每个不完整的红色线段，将其顶点和边（以及紧接在线段下游方向的边）染成灰色，之后将忽略这些部分。

如果在此过程结束时，H 之外的顶点被染成了红色或紫色，我们就称该顶点为着色顶点。请注意，根据构造，$G\backslash H$ 中的所有着色线段都是高度恰好为 ℓ的树，其中最上游顶点与唯一的最下游顶点之间的距离恰好为 $\ell-1$；特别地，所有线段的大小至少为 ℓ。

命题 3.4 将由以下两个引理推出。陈述第一个引理时，我们假设 $p_c<p'<p$。若 $C^{(2)}(C_{\max}(G_n(p')))$ 达到大小 $\epsilon n/8$，则令 H 为其；否则，将 $C^{(2)}_{\max}(G_n(p'))$ 扩展为大小为 $\lfloor\frac{\epsilon n}{8}\rfloor$ 的连通子图 $H\subseteq C_{\max}(G_n(p'))$（鉴于推论 3.2，我们可以假设 that

$C_{\max}(G_n(p'))$ 具有非平凡的 2-核）。我们计算 $C_{\max}(G_n(p'))$ 中顶点相对于 H 的深度，并对 $C_{\max}(G_n(p'))$ 应用参数为 ℓ 的着色方案，该参数将在后文选择。

引理 3.5. 假设 $p > p_c$ 和 $\epsilon < \zeta^{(2,\infty)}(p)$。要使 (10) 成立，只需证明当 $p' < p$ 充分接近 p 且 ℓ 充分大时

$$\frac{1}{n}|C^{(2,\ell)}(C_{\max}(G_n(p'))) \backslash C^{(2)}(C_{\max}(G_n(p)))| \leq \frac{\epsilon}{2} \qquad whp. \tag{11}$$

此外，如果式 (11) 中的事件未发生，则存在一组彩色段 S in $C_{\max}(G_n(p'))$，其合计大小至少为 $\frac{\epsilon n}{8}$ with $S \cap C^{(2)}(G_n(p)) = \emptyset$。

该引理的第一部分很直接：根据 $\zeta^{(2,\infty)}$ 关于 $p \neq p_c$ 的连续性（参见附录 D），我们可以选取 $p' \in (p_c, p)$，使得 $\zeta^{(2,\infty)}(p') \geq \zeta^{(2,\infty)}(p) - \frac{\epsilon}{4}$。当 ℓ 充分大时，从 (10) 推导出 (11) 可由引理 3.3 得出。第二部分的证明涉及对着色方案的详细分析，故留待附录 G.3。

第二个引理将由 G_n 的弱扩张及其有界平均度推出。证明再次使用撒点论证。

引理3.6. 存在取决于 ϵ 的 $\delta > 0$，使得，对于 $p_c < p' < p$ 以及充分大的 ℓ，存在一组着色线段 S，位于 $C_{\max}(G_n(p'))$中，且总大小至少为 $\frac{\epsilon n}{8}$，满足 $S \cap C^{(2)}(G_n(p)) = \emptyset$ 的概率至多为 $2^{n/\ell} \cdot (1-\beta^L)^{\frac{\delta n}{2} - \frac{n}{\ell}}$，其中 $\beta = 1 - \frac{1-p}{1-p'}$ 以及 $L = \bar{d}/\delta$。

证明。总共有 n 个顶点，且每个彩色段的大小至少为 ℓ，因此彩色段至多有 n/ℓ 个，故其子集至多有 $2^{n/\ell}$ 种 S。对于任意一组 S 彩色段，若其总大小为 $\frac{\epsilon n}{8}$ 或更大，则根据弱扩张性质，在这些彩色段与原始图 G_n 中的 H 之间存在至少 δn 条不相交路径，其中某个 $\delta > 0$ 取决于 ϵ；此外，至少一半路径的长度不超过 $L = d/\delta$。令 P 表示此类路径的集合。我们通过喷洒从 $G_n(p')$ 得到 $G_n(p)$，即以概率 β 独立地恢复 $E(G_n)$ 中的每条边；在此过程中，P 中的每条路径被恢复的概率至少为 β^L。

聚焦 S 中的一个特定彩色段 A，它是 $C_{\max}(G_n(p')) \backslash H$ 中的一棵（子）树。令 u 为 A 中最下游的顶点，该顶点是唯一的；令 v 为 u 的直接下游（可能位于 H 中）。如果喷洒恢复了一条连接 A 和 H、且不经过直接下游边 (u,v) 的路径，那么 u 必定处于一个环中，并成为 $C^{(2)}(G_n(p))$ 的一部分：为说明这一点，注意从 $G_n(p)$ 中移除 (u,v) 后，u 和 v 仍然连通——通过新恢复的、通向 H 的路径以及 v 现有的下游。

现在回到 P 中 S 与 H 之间的不相交短路径集合。由于 S 最多包含 n/ℓ 条着色线段，P 中至多有 n/ℓ 条路径经过 S中某条彩色段的直接下游边。对于 $\ell > \frac{2}{\delta}$，P中至少有 $\frac{\delta n}{2} - \frac{n}{\ell}$ 条不相交路径不会经过 S中各线段的直接下游边。如上所述，如果这些路径中的任何一条通过喷洒恢复，那么 S 就不可能完全位于 $C^{(2)}(G_n(p))$ 之外。因此，S 完全位于 $C^{(2)}(G_n(p))$ 之外的概率至多为 $(1-\beta^L)^{\frac{\delta n}{2} - \frac{n}{\ell}}$。对 S的所有可能选择取并集界，可确保存在一个着色线段集合 S，其总大小至少为 $\frac{\epsilon n}{8}$，且仍完全位于 $C^{(2)}(G_n(p))$ 之外的概率至多为 $2^{n/\ell} \cdot (1-\beta^L)^{\frac{\delta n}{2} - \frac{n}{\ell}}$，正如我们所声称的那样。 □

命题 3.4 的证明直接由这两个引理推出，详细内容请参见附录 G.3。将命题 3.4 与匹配的上界（附录 E 中的引理 E.2）结合，可得到巨型 2-核的相对大小依概率收敛。证明整个 2-核大小的相对大小依概率收敛所需的额外分析将在附录 F 中介绍。综合这些结果，即可证明定理 1.4。

4 算法 1 的鲁棒性

最后，我们分析算法 1 在模型设定错误下的鲁棒性。我们简要概述分析过程，定理 1.3 的完整证明见附录 H。

请注意，根据定理 1.4，当 $p \neq p_c$ 时， $C^{(2)}(G_n(p))$ 和 $C^{(2)}_{\max}(G_n(p))$ 的实际分数规模分别可由渗流函数 $\zeta^{(2)}(p)$ 和 $\zeta^{(2,\infty)}(p)$ 很好地近似。因此，只需证明估计值 $I^{(2)}$ 和 $I^{(2,\infty)}$ 在极限下也分别收敛到对应渗流函数的 ϵ-邻域。目前，我们先关注将 $I^{(2,\infty)}$ 作为 $\zeta^{(2,\infty)}(p)$ 的估计；对 $I^{(2)}$ 的分析类似。具体而言，我们将说明如何选择这些参数，由此可直接得出证明。

首先，选择 ℓ 足够大，使得对于所有 $\ell' \geq \ell$ 都有 $|\zeta^{(2,\ell')}(p) - \zeta^{(2,\infty)}(p)| \leq \epsilon/16$。这确保了当 $R_t \geq \ell$ 时 2-核的局部近似质量。当 $R_t < \ell$ 时，该近似可能较差，但这仅在 $|\mathcal{N}_\ell(v_t; G_n)| > K$ 时发生。通过选择 K 足够大，我们可以确保在给定 $G_n(p)$ 的条件下（即相对于 v_t 的随机选择）， $R_t < \ell$ 的概率至多为 $\epsilon/8$ 以高概率（相对于随机的 G_n 和渗流）。总体而言，对 ℓ 和 K 的选择保证了每个 $I_t^{(2,\infty)}$ 在给定 $G_n(p)$ 的条件下，其期望值与 $\zeta^{(2,\infty)}(p)$ 的偏差以高概率至多为 $\epsilon/4$。最后，由于 $I^{(2,\infty)}$ 是 T 个独立同分布的伯努利随机变量 $\{I_t^{(2,\infty)}\}_{t\in[T]}$ 的平均值，它满足经典的 Hoeffding 集中性；通过选择 $T = O\big(\frac{1}{\epsilon^2}\log\frac{1}{\epsilon}\big)$，我们可以确保 $I^{(2,\infty)}$ 与 $\mathbb{E}\big[I_t^{(2,\infty)} | G_n(p)\big]$ 的偏差至多为 $\epsilon/4$，其概率为 $\epsilon/2$。通过对参数作出这些选择，我们保证估计值 $I^{(2,\infty)}$ 与 $\zeta^{(2,\infty)}(p)$ 之间的加性误差在 $\frac{\epsilon}{2}$ 以内。

用于确定 2-核规模的局部算法

我们以伪代码形式给出算法 1 的明确实现。该算法以图G、局部性参数 K 和重复次数 T 作为输入，从$V(G)$中独立同分布地抽取 T 个顶点，对每个样本运行局部探索子程序，并返回T 个样本的平均值。

算法 2 详细说明了使用广度优先搜索的局部探索子程序。它根据局部邻域判断一个顶点是否可能属于 G 的（巨型分量）2-核。

算法 2：用于 2-核候选判定的局部探索子程序

图 $G=(V,E)$，局部性参数 K

输出： $C^{(2)}(G)$ 和 $C^{(2)}(C_{\max}(G))$ 可能属于其中的指示变量

均匀随机抽取样本 $v\in V$

如果 $\deg(v)\geq K$ 那么 // 一个高度顶点

  返回 $I^{(2)}=I^{(2,\infty)}=1$

设 $r=0$, $S=1$ // 初始化探索半径和邻域大小

初始化 $q=$队列(v)

对 $u\in$ 的邻居执行

  设置 $r_u=r+1$、$p_u=u$ // 存储到 v 的距离，以及通向 v 的祖先

  将 $\text{reach}_u=r_u$、$\text{cycle}_u=$ 设为 False

  $S=S+1$

  q.入队(u)

当 $q\neq\emptyset$ 时

  $u=q$.弹出()

  $r=r_u$

  对于 $w\in$邻居(u) do

    如果 $p_w=$ 未定义，则

      集合 $r_w=r+1$, $p_w=p_u$ // 存储到 v 的距离以及朝向 v的祖先

      设置范围 $p_u=r_w$

      $S=S+1$

      如果 $S>K$ 那么

        跳出循环并当

      q.入队(w)

    else

      设置循环$_{p_u}=$真

如果 $S\geq K$ 那么 // 大小至少为 K的连通分量

  返回 $I^{(2)}=I^{(2,\infty)}=\mathbb{I}\{\sum_{u\in\text{邻居}(v)}\mathbb{I}\{\text{环}_u\text{或到达}_u=r\}\geq 2\}$

否则 // 大小小于 K 的连通分量

  返回 $I^{(2)}=\mathbb{I}\{\sum_{u\in\text{邻居}(v)}\mathbb{I}\{\text{环}_u\}\geq 2\}$和 $I^{(2,\infty)}=0$

B 局部收敛的推论

关于局部收敛图序列上的渗流，以下事实将被反复使用。

引理 B.1（引理 3.1, [3]）。设 $\{G_n\}_{n\in\mathbb{N}}$ 是一个局部收敛到 $(G,o)\in\mathcal{G}^*$、其分布为 μ 的弱扩张图序列。设 $f\colon\mathcal{G}^*\to\mathbb{R}$ 是一个有界函数，仅依赖于有限邻域

根周围的有限邻域。那么，对所有 $p\in[0,1]$，

$$\frac{1}{n}\sum_{v\in V(G_n)} f(G_n(p),v)\xrightarrow{p}\mathbb{E}[f(G(p),o)]. \tag{12}$$

下一个引理可由一个推论直接得到。

引理 B.2（有界邻域大小）。设 $\{G_n\}_{n\in\mathbb{N}}$ 是一个局部收敛序列，其极限为 $(G,o)\in\mathcal{G}^*$，且分布为 μ。对于任意 $r\in\mathbb{N}$ 和 $\epsilon\in(0,1)$，存在 $N\in\mathbb{N}$ 使得

$$\frac{1}{n}\sum_{v\in V(G_n)}\mathbb{I}\{|\mathcal{N}_r(v)|\geq N\}\leq\epsilon \qquad whp. \tag{13}$$

证明。令 $f_N(G,o)=\mathbb{I}\{|\mathcal{N}_r(o)|\geq N\}$，这显然是局部且有界的。由引理 B. 1，

$$\frac{1}{n}\sum_{v\in V(G_n)} f_N(G_n,v)\xrightarrow{p}\mathbb{E}[f_N(G,o)]=\mathbb{P}(|\mathcal{N}_r(o)|\geq N).$$

由于 G 几乎必然局部有限，当 N 足够大时，我们有 $\mathbb{P}(|\mathcal{N}_r(o)|\geq N)\leq\epsilon/2$，证毕。

□

C 巨型连通分量的大小

C.1 喷洒与巨型连通分量的大小

定理 2.1 的证明与 [3] 中命题 3.5 的证明类似。回顾 [3]，若 G 是一个（ϵ，α）-大集合扩张图，则

$$\phi(G,\epsilon):=\min_{A\subseteq V,\epsilon n\leq|A|\leq(1-\epsilon)n}\frac{|E(A,V\backslash A)|}{|A|}\geq\alpha.$$

量 ϕ（G，ϵ）称为 G 的 ϵ-大集合扩张。我们对弱扩张图序列的定义只要求，对于任意 $\epsilon>0$（在我们的定义中，$\delta=\epsilon\phi$（G，ϵ））， ϵ-大集合扩张渐近地有一个严格大于零的下界。

下面的引理重述了 [3] 中的引理 3.3，并概括了喷洒论证的要旨。为便于查阅，我们在此给出证明；该证明与 [3] 中的证明完全相同。

引理 C.1（喷洒，引理 3.3, [3]）。设 $G=(V,E)$ 是一个具有 n 个顶点的（ϵ，α）-扩张图，且其平均度至多为 d。取实常数 $\kappa>0$ 和 β、$\epsilon\in$（0，1）。设 S 是 G 的一族互不相交的连通子图，且每个子图的大小至少为 κ。考虑这样的事件：存在将 S 划分为 S_0 和 $S\backslash S_0$ 的一种划分，其中二者的大小均至少为 ϵn，且它们在 G 中不连通（β）至多为 $\exp\left(\frac{n}{\kappa}-\beta^{\bar{d}/\alpha\epsilon}\frac{\alpha\epsilon n}{2}\right)$。

证明。只需考虑由 S 中子图的若干完整部分组成的 S_0；否则 S_0 和 $S\backslash S_0$ 显然是连通的。我们希望给出连接 S_0 中的顶点与 $S\backslash S_0$ 中顶点的（不相交）路径数量下界。这是两个顶点子集之间的最大流，等于最小割。根据 G 的 ϵ-大集合扩张性质，在 $|S_0|$、$|S\backslash S_0|\geq\epsilon n$ 的条件下，最小割至少为 $\alpha\epsilon n$，因此两组之间至少存在 $\alpha\epsilon n$ 条不相交路径。由于边的总数不超过 $nd/2$，至少一半的路径长度有上界 $\frac{\bar{d}}{\alpha\epsilon}$。因此，在 G（β）中， S_0 与 $S\backslash S_0$ 不连通的概率至多为 $(1-\beta^{\bar{d}/\alpha\epsilon})^{\alpha\epsilon n/2}$。 S_0 至多有 $2^{n/\kappa}$ 种可能的取值，因此总体概率的上界为 $2^{n/\kappa}(1-\beta^{\bar{d}/\alpha\epsilon})^{\alpha\epsilon n/2}\leq\exp(n/\kappa-\beta^{\bar{d}/\alpha\epsilon}\alpha\epsilon n/2)$，证毕。

□

注意，喷洒论证是针对任意固定的 $\epsilon>0$ 构造的。因此，在证明巨型连通分量的唯一性和局部性时，只需考虑一个固定的 $\epsilon>0$，并使用 ϵ-大集合扩张性质（它取决于 ϵ）。基于此，我们考虑一个局部收敛的弱扩张图序列$\{G_n\}_{n\in\mathbb{N}}$。与 [3] 中完全相同的分析同样适用，并得出以下命题。

命题 C.2. 对任意 $\epsilon>0$ 以及 ζ 的任意连续点 $p\in[0,1]$，

$$\mathbb{P}\Big(\Big|\frac{1}{n}|C_{\max}(G_n(p))|-\zeta(p)\Big|\geq\epsilon\Big)\to 0. \tag{14}$$

C.2 渗流函数 ζ的连续性

由命题 C.2 可知，要得到定理 2.1 中所述的 $C_{\max}(G_n(p))$ 相对大小的收敛，只需证明渗流函数 ζ 的连续性；而这几乎可以直接由 [3] 中的分析得出。将大集合扩张稍作推广为弱扩张，并不会影响证明，唯一需要特别注意的是如何选取扩张参数。

首先，我们注意到弱扩张图序列的局部极限是遍历的。

引理 C.3（参见引理 A.1, [3]）。设 $\{G_n\}_{n\in\mathbb{N}}$ 是一个（可能为随机的）有界平均度弱扩张图序列，且局部收敛到 $(G,o)\in\mathcal{G}^*$，其分布为 μ。那么 μ 是遍历的，因此在 $\mathcal{G}^*$ 上的单模概率测度中是极端的。

证明概要。除以下细微差异外，证明几乎与 [3], 中引理 A.1 的证明完全相同：一旦固定 $\epsilon\leq p_0/8$，我们的图序列的 ϵ-大集合扩张渐近为 $\alpha=\alpha(\epsilon)$，而不是像原证明中那样为某个绝对量（与 ϵ 无关）。然而，这对我们并无影响，因为证明的其余部分依赖于 ϵ 和 α 的固定取值。

□

特别地，如果 μ 是弱扩张图序列的局部极限，那么在 $(G,o)\sim\mu$ 上的渗流临界概率几乎必然为常数 μ。令 $p_c=p_c(\mu)$ 表示该临界概率，即，

$$\inf\{p:\mathbb{P}(|C(o;G(p))|=\infty|(G,o))>0\}=p_c \qquad \text{a.s.} \tag{15}$$

根据 [1], 中的定理 6.7，对于满足 $p_c<p_1<p_2\leq 1$ 的所有 p_1、p_2，$G(p_2)$ 中的每个无限簇都包含 $G(p_1)$ 中的一个无限簇（在渗流的标准耦合下）μ-几乎必然成立。这使我们能够立即得出结论：ζ 在所有 $p\neq p_c$ 处连续。

命题 C.4（ζ的连续性。参见推论 2.2, [3]）。设 μ 是一个弱扩张图序列的局部极限，$\zeta:[0,1]\to[0,1]$ 是相应的渗流函数。那么，ζ 在除 $p=p_c$ 以外的所有 $p\in[0,1]$ 处连续。

定理 2.1 现可由命题 C.2 和 C.4 立即得出。此外，稍复杂一些的分析表明，对于所有 $p\in[0,1]$（包括 $p=p_c$），第二大连通分量 $C_2(G_n(p))$ 的规模以高概率为次线性的（参见 [3, 附录 C]）。

以下推论直接由 $|n^{-1}C_{\max}(G_n(p))|$ 收敛到 $\zeta(p)$ 得出。

推论 C.5. 对于任意 $p\neq p_c$ 和任意 $\epsilon>0$，存在足够大的 K，使得

$$\frac{1}{n}\sum_{v\in V(G_n)}\mathbb{I}\{|C(v;G_n(p))|\geq K, v\notin C_{\max}(G_n(p))\}\leq\epsilon \qquad whp. \tag{16}$$

Proof. The inequality is only interesting when $|C_{\max}(G_n(p))| \geq K$ – otherwise the left hand side is zero.
当 $|C_{\max} G_n p | \geq K$ $K\}-\mathbb{I}\{v \in C_{\max} G_n p \}$ $\mathbb{I}\{|C v G_n p | \geq K v \notin C_{\max} G_n p \} = \mathbb{I}\{|C v G_n p | \geq$ (( )) 时，我们有 (; ( ))、(( )) (; ( )) $\zeta_{\geq K} p = \mathbb{P}_{(G,o)\sim\mu} |C o G p|\geq$ (( ))。选择足够大的，使得 ( ) : ((; ( ))K $=$) $\leq \zeta(p) + \epsilon/2$，这可由收敛性 $\zeta(p) \lim_{K\to\infty} \zeta_{\geq K}(p)$ 保证（例如，单调收敛）。根据定理 2.1,

$$\frac{1}{n}\sum_{v\in V(G_n)} \mathbb{I}\{v \in C_{\max}(G_n(p))\} = \frac{1}{n}|C_{\max}(G_n(p))| \xrightarrow{p} \zeta(p),$$

并且根据引理 B.1,

$$\frac{1}{n}\sum_{v\in V(G_n)} \mathbb{I}\{|C(v;G_n(p))| \geq K\} \xrightarrow{p} \zeta_{\geq K}(p).$$

取二者之差即可得到该推论。 □

D 2-核渗流函数的连续性

我们现在研究弱扩张图序列的局部极限中函数 $\zeta^{(2)}$ 和 $\zeta^{(2,\infty)}$ 的连续性。证明思路与 [3] 中渗流函数 ζ 连续性的证明非常相似（见第 2.3 节和附录 B）。对于 $\zeta^{(2,\infty)}$，我们将像 [43] 中那样分析渗流图中的无限连通分量。一旦证明了 $\zeta^{(2,\infty)}$ 的连续性，便很容易推出 $\zeta^{(2)}$ 的连续性。本节中，我们将关注随机有根图 $(G,o)\in\mathcal{G}^*$，其分布为 μ，并假设它是一个弱扩张图序列的局部极限。

引理 D.1. T 函数 $\zeta^{(2,\infty)}$与 ζ 一样，在 $[0,1]$ 上处处连续，除非在 p 处可能不连续 c.

证明。由于 p_c $\zeta^{(2,\infty)}(p) \equiv 0$ 对 $p < p_c$成立，故下方的连续性显然。考虑任意 $p > p_c$。由命题 3.1, $\zeta^{(2,\infty)}(p) > 0$。根据定义,

$$\zeta^{(2,\infty)}(p) = \mathbb{P}\big(o \in C^{(2,\infty)}(G(p))\big). \tag{17}$$

不难验证， $o \in C^{(2,\infty)}(G(p))$当且仅当 $\tilde{H} = C(o;G(p))\backslash\{o\}$ 满足以下两个条件之一:

˜ ˜(I) 在 $G(p)$ 中， o 至少有两个邻居 u 和 v，使得 $C(u;H)$ 和 $C(v;H)$（可能属于同一个连通分量）都为无限;

˜(II)在 $G(p)$中， o 的恰好一个邻居 u 满足 $|C(u;H)| = \infty$，并且 o至少还有另一个邻居 v满足 $C(o;H)$包含一个环或包含 o的另一个邻居 $w \neq u$、v。

˜根据 [1], 中的定理 6.7， $G(p)\backslash\{o\}$ 中的每个无限连通分量（尤其是 H中的每个无限连通分量）几乎必然都包含 $G(p')\backslash\{o\}$ 中的一个无限连通分量，其中 $p' \in (p_c,p)$：否则，以正概率存在 $G(p)\backslash\{o\}$中的一个无限连通分量，它通过 $k \in \mathbb{N}$ 条边与 o 在 $G(p)$中相连，却不包含 $G(p')$中的任何无限连通分量；通过对这些 k 条边上的边渗流进行重新采样， $G(p)$就有正概率包含一个无限连通分量，而其中没有任何无限子连通分量属于 $G(p')$，这与 [1]中的定理 6.7矛盾。

现在只需证明，在给定 $o \in C^{(2,\infty)}(G(p))$ 的条件下（这是成立的，因为此类事件发生的概率严格为正），我们几乎必然有 $o \in \bigcup_{p''<p} C^{(2,\infty)}(G(p''))$。这等价于

$$\mathbb{P}\left(o \in \bigcup_{p''<p} C^{(2,\infty)}(G(p'')) \,\middle|\, o \in C^{(2,\infty)}(G(p))\right) = \sup_{p''<p} \frac{\zeta^{(2,\infty)}(p'')}{\zeta^{(2,\infty)}(p)} = 1, \tag{18}$$

这立即表明在 p 处左连续；结合 $\zeta^{(2,\infty)}$ 的右连续性，即可得到所需结果。

为证明上述结论，令 $H = C(o;G(p))$，并以 $o \in C^{(2,\infty)}(G(p))$ 和 H 的实现为条件。注意，情形 (I) 和 (II) 互不相容。先考虑情形 (I)。选取 $p' \in (p_c,p)$。不妨假设 $C(u;H)$ 和 $C(v;H)$ 各自包含 $G(p')$ 的一个无限连通分量，因此在 $G(p)$ 中，存在分别从 u、v 通往它们的有限路径。将 $p'' < p$ 取为充分接近 p，这些路径连同边 (o,u) 和 (o,v) 最终都会包含在 $G(p'')$ 中，这表明 $o \in C^{(2,\infty)}(G(p''))$。对于情形 (II)，关于 u 的推理保持不变，并且在 $G(p)$ 中，存在一条从 v 通往某个（有限）环或 w 的有限路径。当 $p'' < p$ 充分接近 p 时，这条路径同样会保留在 $G(p'')$ 中，这表明 $o \in C^{(2,\infty)}(G(p''))$。这就证明了该结论，并完成了证明。 □

推论 D.2。 函数 $\zeta^{(2)}$ 与 ζ 一样，在 $[0,1]$ 上处处连续，除非在 p_c 处 .

证明。为研究 $\zeta^{(2)}$ 的连续性，我们考虑它与 $\zeta^{(2,\infty)}$ 的逐点差，记为 $\zeta^{(2,<\infty)}(p) := \zeta^{(2)}(p) - \zeta^{(2,\infty)}(p) = \mathbb{P}(o \in C^{(2,\infty)}(G(p)) \ |C(o;G(p))| < \infty)$。由于我们不处于有限连通分量中，显然，在给定 $C(o;G(p))$ 的实现条件下，当 $p'' < p$ 足够接近 p 时，$C(o;G(p))$ 中不会有边消失，因此 $o \in C^{(2,<\infty)}(G(p''))$ 与 $|C(o;G(p''))| < \infty$。因此，$\zeta^{(2,<\infty)}$ 在整个单位区间上左连续。右连续性也显而易见。因此，$\zeta^{(2,<\infty)}$ 连续，而 $\zeta^{(2)} = \zeta^{(2,<\infty)} + \zeta^{(2,\infty)}$ 在 $[0,1]$ 上连续，除非在 p_c 处。 □

E 上界

本节中，我们将使用标准技术（除了已知巨型连通分量外，不依赖任何扩张假设），证明 $C^{(2)}(C_{\max}(G_n(p)))$ 相对大小的一个高概率上界。结合命题 3.4，即可得到巨型 2-核相对大小依概率收敛，也就是以下定理。

定理 E.1. 对任意 $\epsilon > 0$ 和 $p \neq p_c$，$\frac{1}{n}|C^{(2)}_{\max}(G_n(p))| \xrightarrow{p} \zeta^{(2,\infty)}(p)$。当 $p = p_c$ 时，右侧构成一个高概率上界。

整个图 $G_n(p)$ 的 2-核也有类似结果，详见附录 F。本附录中，我们证明以下引理。

引理 E.2. 对任意 $\epsilon > 0$ 和 $p \in (0,1],$，我们有

$$\frac{1}{n}|C^{(2)}(C_{\max}(G_n(p)))| \leq \zeta^{(2,\infty)}(p) + \epsilon \qquad whp. \tag{19}$$

证明。注意，当 $\frac{1}{n}|C_{\max}(G_n(p))| < \epsilon$ 时，这个引理显然成立。因此，我们可以假设 $|C_{\max}(G_n(p))| \geq \epsilon n$。

回顾第 3.1 节中对有根图 (G,o) 的局部事件 $o \in C^{(2,\ell)}(G)$ 的定义；特别地，我们有 $C^{(2,\infty)}(G) = \bigcap_{\ell \in \mathbb{N}} C^{(2,\ell)}(G)$。令 $\zeta^{(2,\ell)}(p) = \mathbb{P}_{(G,o)}(o \in C^{(2,\ell)}(G(p)))$，它从上方趋近于 $\zeta^{(2,\infty)}(p) = \mathbb{P}_{(G,o)}(o \in C^{(2,\infty)}(G(p)))$。通过选取足够大的 ℓ s，我们可以保证 $\zeta^{(2,\ell)}(p) \leq \zeta^{(2,\infty)}(p) + \epsilon/3$。

由于 $C^{(2,\ell)}$ 描述的是一个局部事件，根据引理 B.1，G_n 的局部收敛意味着

$$\frac{1}{n}\sum_{v \in V(G_n)} \mathbb{I}\{v \in C^{(2,\ell)}(G_n(p))\} = \frac{1}{n}|C^{(2,\ell)}(G_n(p))| \xrightarrow{p} \zeta^{(2,\ell)}(p). \tag{20}$$

特别是，

$$\frac{1}{n}|C^{(2,\ell)}(G_n(p))| \leq \zeta^{(2,\ell)}(p) + \frac{\epsilon}{3} \leq \zeta^{(2,\infty)}(p) + \frac{2\epsilon}{3} \qquad \text{whp.} \tag{21}$$

现在只需对 $C^{(2)}(C_{\max}(G_n(p)))\backslash C^{(2,\ell)}(G_n(p))$进行界定。注意，属于 $C^{(2)}(G_n(p))$但不属于 $C^{(2,\ell)}(G_n(p))$的顶点，其 ℓ-邻域必然包含$G_n(p)$的一个完整连通分量。特别是， $v\in C^{(2)}(C_{\max}(G_n(p)))\backslash C^{(2,\ell)}(G_n(p))$意味着 $|\mathcal{N}_\ell(v;G_n)|\geq |C_{\max}(G_n(p))|\geq \epsilon n$。由引理 B.2，我们得出

$$\frac{1}{n}|C^{(2)}(C_{\max}(G_n(p)))\backslash C^{(2,\ell)}(G_n(p))|\leq \frac{1}{n}\sum_{v\in V(G_n)}\mathbb{I}\{|\mathcal{N}_\ell(v)|\geq \epsilon n\}\leq \frac{\epsilon}{3} \qquad \text{whp.} \tag{22}$$

结合式 (21) 即完成证明。 □

F 整个 2-核的大小

现在我们考察 $G_n(p)$ 的整个 2-核的大小，并证明以下定理。

定理 F.1. 对任意 $\epsilon>0$ 和 $p\neq p_c$， $\frac{1}{n}|C^{(2)}(G_n(p))|\xrightarrow{p}\zeta^{(2)}(p)$。当 $p=p_c$ 时，右侧是一个高概率上界。

由于我们已知，根据定理 E.1，当 $p\neq p_c$ 时，相对大小 $C^{(2)}(C_{\max}(G_n(p)))$ 依概率收敛到 $\zeta^{(2,\infty)}(p)$，因此只需关注量 $C^{(2)}(G_n(p)\backslash C_{\max}G_n(p)))$，也就是 $G_n(p)$ 中最大连通分量之外的 2-核。

只需证明以下结果，由此即可推出定理 F.1。

命题 F.2。对于任意 $p\in[0,1]$,

$$\frac{1}{n}|C^{(2)}(G_n(p)\backslash C_{\max}(G_n(p)))|\xrightarrow{p}\zeta^{(2)}(p)-\zeta^{(2,\infty)}(p). \tag{23}$$

证明。给定一个图 G，令 $\zeta^{(2,<\infty)}(p)=\zeta^{(2)}(p)-\zeta^{(2,\infty)}(p)$，并令 $Z_{<\ell}(G)$（分别地， $Z_{\geq\ell}(G)$）表示 G 的连通分量中规模小于 ℓ（分别地，不小于 ℓ）的顶点集合。

固定 ome $\epsilon>0$。根据推论 C.5，存在一个充分大的常数 ℓ，使得

$$\frac{1}{n}|Z_{\geq\ell}(G_n(p))\backslash C_{\max}(G_n(p))|\leq\frac{\epsilon}{3} \qquad \text{whp.} \tag{24}$$

注意，当 $|C_{\max}(G_n(p))|\geq\ell$ 时，我们有

$$G_n(p)\backslash C_{\max}(G_n(p))=Z_{<\ell}(G_n(p))\sqcup\big(Z_{\geq\ell}(G_n(p))\backslash C_{\max}(G_n(p))\big)$$

其中 $\sqcup$ 表示不交并；当 $|C_{\max}(G_n(p))|<\ell$ 时， $C^{(2)}(G_n(p)\backslash C_{\max}(G_n(p)))$ 与 $C^{(2)}(Z_{<\ell}(G_n(p)))$ 至多相差 ℓ 个顶点。无论哪种情况，只需证明：通过将 ℓ 选得足够大，我们可以保证

$$\left|\frac{1}{n}|C^{(2)}(Z_{<\ell}(G_n(p)))|-\zeta^{(2,<\infty)}(p)\right|\leq\frac{2\epsilon}{3} \qquad \text{whp.} \tag{25}$$

定义 $\zeta^{(2,<\ell)}(p):=\mathbb{P}_{(G,o)\sim\mu}(o\in C^{(2)}(G(p))\ |C(o;G(p))|<\ell)$。注意，相应事件构成一个单调递增序列，因为 $\ell\to\infty$，因此，当 $(G,o)\sim\mu$ 时，

$$\lim_{\ell\to\infty}\zeta^{(2,<\ell)}(p)=\lim_{\ell\to\infty}\mathbb{P}(o\in C^{(2)}(G(p))\ |C(o;G(p))|<\ell)=\mathbb{P}(o\in C^{(2)}(G(p))\ |C(o;G(p))|<\infty)=\zeta^{(2,<\infty)}(p).$$

取足够大的 ℓ，即可保证 $|\zeta^{(2,<\ell)}(p)-\zeta^{(2,<\infty)}(p)|\leq\epsilon/3$。此外，对于有根图 (G,o)，事件 $o\in C^{(2)}(G)$ 且 $|C(o;G)|<\ell$ 是局部的，根据引理 B.1,

$$\frac{1}{n}|C^{(2)}(Z_{<\ell}(G_n(p)))|=\frac{1}{n}\sum_{v\in V(G_n)}\mathbb{I}\{v\in C^{(2)}(G_n(p)),|C(v;G_n(p))|<\ell\}\xrightarrow{p}\zeta^{(2,<\ell)}(p). \tag{26}$$

因此，式（25）成立，我们的证明也就完整了。 □

G 第3节的证明细节

G.1 第3.1节的证明细节

我们首先证明一个一般引理：利用根的邻域中任意顶点对应的局部事件的概率，给出根对应局部事件概率的下界；这实际上表明，前一个引理对任意选择的 $r \in \mathbb{N}$ 都成立。这直接源于局部收敛（依概率）和极限的次模性。

引理 G.1。设 $\mathcal{E}(G,o)$ 是一个局部事件。那么对于任意 $\delta > 0$、$r \in \mathbb{N}$ 和 $p \in [0,1]$，存在一个常数 $D \in \mathbb{N}$（取决于 δ 和 r），使得

$$\mathbb{P}(\exists u \in \mathcal{N}_r(o;G) \ s.t. \ \mathcal{E}(G(p),u)) \leq D\mathbb{P}(\mathcal{E}(G(p),o)) + \delta. \tag{27}$$

其中所有概率均是针对 $(G,o) \sim \mu$ 和渗流的随机性而言。

证明。注意，只需对 $r = 1$ 证明 (27)；对于一般情形，只需基于以下事实使用归纳法：$\mathbb{P}(\exists u \in \mathcal{N}_r(o;G) \text{ s.t. } \mathcal{E}(G(p)\ u)) = \mathbb{P}(\exists v \in \mathcal{N}_1(o;G)\ u \in \mathcal{N}_{r-1}(v;G) \text{ s.t. } \mathcal{E}(G(p)\ u))$。由于 μ 是局部有限图空间上的一个分布，我们有

$$\deg(o;G) \vee \max_{u \sim_G o} \deg(u;G) < \infty \qquad \text{a.s.},$$

其中，记号 $u \sim v$ 表示两个顶点 u、$v \in V$ 的邻接关系（在 G 中）。因此，对任意 $\delta > 0$，存在 $D < \infty$，使得 $\deg(o) < D$，且 $\max_{u \sim o} \deg(u;G) < D$ 的概率至少为 $1 - \delta$。

此外，μ 作为局部极限是单模的，因此也是对合不变的 [1, 命题 2.2]。特别地，这意味着

$$\mathbb{E}\left[\sum_{u \sim o} \mathbb{P}(\mathcal{E}(G(p),o), \deg(o) < D | (G,o))\right] = \mathbb{E}\left[\sum_{u \sim o} \mathbb{P}(\mathcal{E}(G(p),u), \deg(u) < D | (G,o))\right]. \tag{28}$$

（注意，这依赖于这样一个事实：$\mathbb{P}(\mathcal{E}(G(p)\ o), \deg(o) < D | (G,o))$ 是一个局部函数。）右侧是 $\mathbb{P}(\exists u \sim o \text{ s.t. } \mathcal{E}(G,u), \deg(u) < D)$ 的上界。因此，

$$\begin{aligned}\mathbb{P}\big(\exists u \sim o \text{ s.t. } \mathcal{E}(G(p),u)\big) &\leq \mathbb{P}(\exists u \sim o \text{ s.t. } \deg(u) \geq D) + \mathbb{E}[\deg(o) \cdot \mathbb{P}(\mathcal{E}(G(p),o), \deg(o) < D | (G,o))] \\ &\leq \delta + D\mathbb{P}(\mathcal{E}(G(p),o))\end{aligned}$$

正如我们所声称的。 □

引理 G.1 基本上使我们能够将 o 属于 $C^{(2)}(G(p))$ 中某个无限连通分支的概率，界定为它位于 $C^{(2)}(G(p))$ 的某个无限连通分支的有界邻域内的概率。为对此作出形式化表述，我们将顶点集 A 的 r-邻域 $\mathcal{N}_r(A;G)$ 定义为由距 A 不超过 r 的顶点所诱导的子图。

推论 G.2. 对于 $p > p_c$ 和任意 $r \in \mathbb{N}$，概率 $\mathbb{P}(o \in C^{(2,\infty)}(G(p)))$ 非零，当且仅当 $\mathbb{P}(o \in \mathcal{N}_r(C^{(2,\infty)}(G(p))))$ 非零。

证明。“仅当”方向显然。对于“若”方向，固定 $p > p_c$ 并假设 $a := \mathbb{P}(o \in \mathcal{N}_r(C^{(2,\infty)}(G(p)))) > 0$。由于对于任意 $\ell \in \mathbb{N}$，a.s. 有 $C^{(2,\infty)}(G(p)) \subseteq C^{(2,\ell)}(G(p))$，因此我们有

$$\mathbb{P}(o \in \mathcal{N}_r(C^{(2,\ell)}(G(p)))) \geq \mathbb{P}(o \in \mathcal{N}_r(C^{(2,\infty)}(G(p)))) = a. \tag{29}$$

根据引理 G.1，存在 $D\in\mathbb{N}$ 使得

$$\mathbb{P}(o\in C^{(2,\ell)}(G(p)))\geq\frac{1}{D}\Big(\mathbb{P}(o\in\mathcal{N}_r(C^{(2,\ell)}(G(p))))-\frac{a}{2}\Big)\geq\frac{a}{2D}. \tag{30}$$

取 $\ell\to\infty$ 的极限，得到 $\zeta^{(2,\infty)}(p)=\mathbb{P}(o\in C^{(2,\infty)}(G(p)))=\lim_{\ell\to\infty}\mathbb{P}(o\in C^{(2,\ell)}(G(p)))\geq\frac{a}{2D}>0$。 □

现在证明，对于 $(G,o)\sim\mu$，o 位于 $C^{(2,\infty)}(G(p))$ 的某个有限邻域内的概率严格为正。

引理 G.3。对于任意 $p>p_c$，存在某个 $r\in\mathbb{N}$，使得 $\mathbb{P}\big(o\in\mathcal{N}_r(C^{(2,\infty)}(G(p)))\big)>0$。

证明。只需证明

$$\mathbb{P}\bigg(o\in\bigcup_{r\in\mathbb{N}}\mathcal{N}_r(C^{(2,\infty)}(G(p)))\bigg)=\mathbb{P}(|C(o;G(p))|=\infty), \tag{31}$$

根据超临界性假设，该值为正。引理由以下事实得出：上述量是 $\mathbb{P}(o\in\mathcal{N}_r(C^{(2,\infty)}(G(p))))$ 的单调极限，其中 $r\to\infty$。

如果 $|C(o;G(p))|=\infty$，但对于所有 $r\in\mathbb{N}$，$o\notin\mathcal{N}_r\big(C^{(2,\infty)}(G(p))\big)$，那么这个无限连通分量的 2-核必为空，因此存在一条从 o 延伸至无穷远的唯一射线。令 $\mathcal{E}_p$ 表示这一事件。(31) 可归结为断言 $\mathbb{P}(\mathcal{E}_p)=0$。

假设 $\mathbb{P}(\mathcal{E}_p)=\epsilon>0$。选择 $p'\in(p_c,p)$ 使得 $\zeta(p')>\zeta(p)-\epsilon$，这得益于 ζ 在 p_c 上方的连续性（见命题 C.4）。通过保留 $G(p)$ 和 $G(p'/p)$ 中的公共边，得到 $G(p')$；等价地，先得到 $G(p)$，然后以概率 p'/p 在其上进行渗流。在给定 $\mathcal{E}_p$的条件下，从 $G(p)$ 中的 o 到无穷远存在唯一一条射线，沿着该射线移除任意一条边都会使 o 处于一个有限连通分量中，因此，以概率 1，o 在 $G(p')$ 中处于一个有限连通分量中。因此，$\zeta(p')\leq\zeta(p)-\epsilon$，矛盾。 □

命题 3.1 的证明现在只需 3 行即可完成。

命题3.1 的证明。第二部分显然，因为 $\zeta(p)=0$ 对于 $p<p_c$。因此假设 $p>p_c$。根据引理 G.3，存在某些 $r\in\mathbb{N}$ 使得 $\mathbb{P}\big(o\in\mathcal{N}_r(C^{(2,\infty)}(G(p)))\big)>0$，而根据推论G.2，这意味着 $\zeta^{(2,\infty)}(p)=\mathbb{P}(o\in C^{(2,\infty)}(G(p)))>0$。 □

最后，我们给出引理 3.3 的省略证明。我们先给出以下引理，它由 $G_n(p)$ 在极限下局部收敛到渗流（见引理 B.1），以及 $C^{(2,\ell)}$ 可以通过局部信息确定这一事实得出。

引理 G.4. 对任意 $p\in[0,1]$ 和任意 $\ell\in\mathbb{N}$，$\frac{1}{n}|C^{(2,\ell)}(G_n(p))|\xrightarrow{p}\zeta^{(2,\ell)}(p):=\mathbb{P}_{(G,o)\sim\mu}(o\in C^{(2,\ell)}(G(p)))$.

N接下来我们证明引理 3.3，该引理曾在第 3.1 节中用于证明命题 3.2。

引理 3.3 的证明。注意到 $C^{(2,\ell)}(G(p))\supseteq C^{(2,\infty)}(G(p))$。因此，$\zeta^{(2,\ell)}(p)\geq\zeta^{(2,\infty)}(p)$，根据命题 3.1，$p>p_c$ 为正。由引理 G.4 可得，$\frac{1}{n}|C^{(2,\ell)}(G_n(p))|\xrightarrow{p}\zeta^{(2,\ell)}(p)$。只需证明，当 ℓ充分大时，

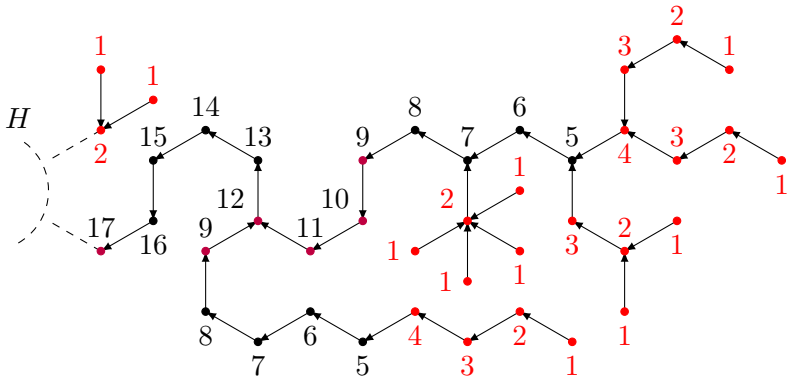
$$\frac{1}{n}|C^{(2,\ell)}(C_{\max}(G_n(p)))|\geq\frac{1}{n}|C^{(2,\ell)}(G_n(p))|-\epsilon\qquad\text{whp}, \tag{32}$$

或等价地,

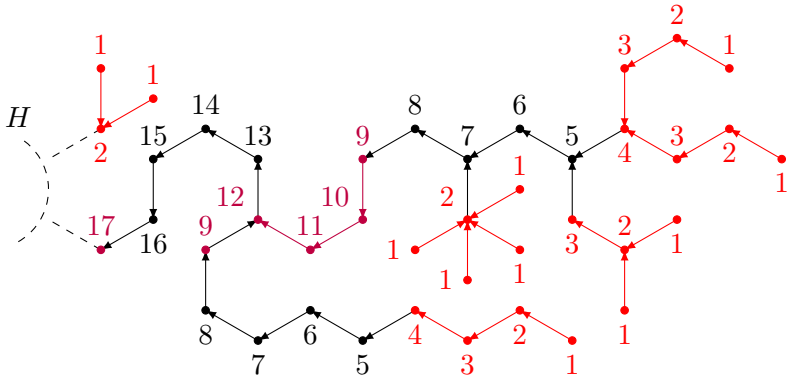
$$\frac{1}{n}\left|C^{(2,\ell)}\big(G_n(p)\backslash C_{\max}(G_n(p))\big)\right| \le \epsilon \qquad \text{whp}, \tag{33}$$

注意， v in $C^{(2,\ell)}\big(G_n(p)\backslash C_{\max}(G_n(p))\big)$ 中的任意顶点都必须属于一个大小至少为 ℓ 、但并非最大连通分量的连通分量。根据推论 C.5，当 ℓ 充分大时，式 (33) 成立。 □

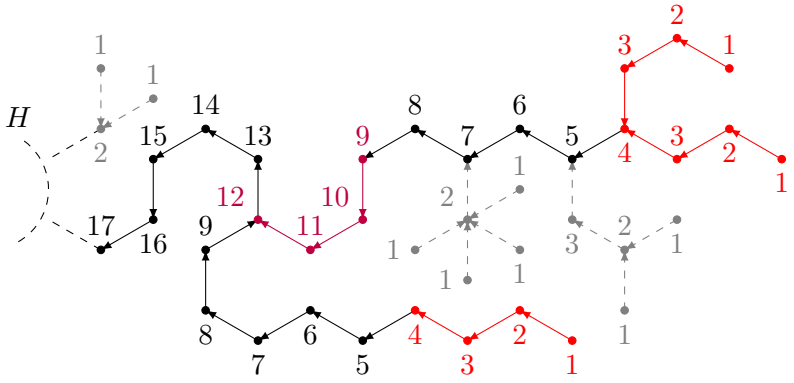
G.2 第3.2节示意图



(a) 根据顶点深度对 2ℓ 取模的结果，将顶点着为紫色或红色。



(b) 将深度连续且颜色相同的顶点之间的边着色，从而形成线段 .



(c) 撤销不完整线段中顶点和边的着色。

图2：将着色方案应用于 $G\backslash H$，并选择 $\ell = 4$。顶点标注了其深度。灰色线段以虚线边表示，说明在后续的撒点论证中会忽略这些线段。请注意，每个红色或紫色线段的大小至少为 ℓ，且对于任意非灰色顶点，完全包含在其（含自身的）上游中的着色线段至少占其常规（含自身）上游的 $1/3$ （见引理 G.5）。显然，假设 $H \supseteq C^{(2)}(G) \neq \emptyset$，在喷洒过程中，任何顶点一旦进入某个环或环间路径，其整个下游都会被纳入 2-核。

G.3 第3.2节的证明细节

本附录详细补充第3.2节中概述巨型 2-核规模上界证明时略去的证明细节。

我们还需要引入一些记号。回顾第3.2节中的着色方案，特别是灰色线段的概念。我们在撒点论证中没有纳入这些线段，因为灰色线段可能很短，从而给并集界带来问题；并集界所依赖的条件是所有线段的长度至少为 ℓ，而彩色段满足这一条件。

如果一个线段不是灰色的，我们就称其为规则线段。回顾命题 3.2 的证明中对顶点的上游和下游的定义；对于顶点 u，我们称 $\mathsf{us}(u) := \{u\} \cup \mathsf{us}(u)$ 为包含自身的上游。顶点 $v \notin H$ 处包含自身的规则上游（记为 $\mathsf{us}^{\mathsf{reg}}(v)$）指 $\mathsf{us}(v)$ 中所有规则顶点的集合； v 处包含自身的彩色上游规模（记为 $\mathsf{us}^{\mathsf{col}}(v)$）是 $G \setminus H$ 中完全包含于 $\mathsf{us}(v)$ 的彩色线段的并集。下一个引理将彩色上游的规模与规则上游的规模联系起来。在陈述引理之前，我们注意到，在 $G \setminus H$ 中，深度为 ℓ 或更小的任何顶点 v 都不能属于 $C^{(2,\infty)}(\mathcal{N}_\ell^+(v))$。这表明， $C^{(2,\ell)}(G) \setminus H$ 中的任何顶点深度都至少为 $\ell+1$，因而必为黑色或紫色。图2对此作了示意。

引理 G.5. 对于任意正则顶点 v, $|\mathsf{us}^{\mathsf{col}}(v)| \geq \frac{1}{3}|\mathsf{us}^{\mathsf{reg}}(v)|$。

注 G.6. 注意， $\mathsf{us}^{\mathsf{col}}(v)$ 是通过从 $\mathsf{us}^{\mathsf{reg}}(v)$ 中移除黑色顶点，以及位于未被 $\mathsf{us}^{\mathsf{reg}}(v)$ 完整包含的片段中的着色顶点而得到的。该引理对这两种影响都给出了界。

证明。 容易看出，当 v 的正则上游是一条长度为 $b = \mathsf{us}^{\mathsf{reg}}(v) - 1 \geq \ell$ 的单一路径时，结论成立：最坏情况出现在 $b = 3\ell - 1$ 的情形，此时 $\mathsf{us}^{\mathsf{col}}(v) = \ell$ （这是由于叶节点处的红色 片段所致，因为下一个 紫色 片段不完整），此时比值为 $\frac{\ell}{3\ell-1} \geq \frac{1}{3}$。

当 v 的正则上游不是一条单一路径时，我们可以将其分解为茎路径，即沿着深度连续的顶点延伸的路径，使用以下过程。

- 初始化 $U = \mathsf{us}_{\mathsf{reg}}(v)$，并继承原始图中的深度编号和颜色。
- 当 U 非空时，迭代地在 U 中找出深度（继承自 G）最大的顶点 u；必要时可任意打破平局。

  – 若 u 为红色，则终止： H 中剩余的所有顶点都必须是红色。– 找一条通向 u的茎路径；若存在多条，则任选其一。[4] 从 U中提取并移除该路径。此茎路径的长度等于 $\mathsf{depth}(u)$ -1，且至少为 ℓ，因为 u 不是红色。

上述过程会将 v 的上游分解为多个长度至少为 ℓ 的茎上游，以及一些额外的红色顶点。着色比例的下界对每个茎上游都成立（其中我们只统计每条路径上相对完整的着色线段）；将这些路径重新组合回 v 的上游，只会增加相对完整的着色线段数量。 □

引理 3.5 的证明。由 $\zeta^{(2,\infty)}$ 关于 $p \neq p_c$ 的连续性（我们已在附录 D 中证明），可选取 $p' \in (p_c, p)$, 使得 $\zeta^{(2,\infty)}(p') \geq \zeta^{(2,\infty)}(p) - \frac{\epsilon}{4}$。当 ℓ 足够大时，推论 3.3 保证

$$\frac{1}{n}|C^{(2,\ell)}(C_{\max}(G_n(p')))| \geq \zeta^{(2,\infty)}(p') - \frac{\epsilon}{4} \geq \zeta^{(2,\infty)}(p) - \frac{\epsilon}{2} \qquad \text{whp}, \tag{34}$$

由此可得从（10）到（11）的化简。

[4]这始终是可行的：如果 u 尚未被移除，那么它的任何直接上游节点 w也一定尚未被移除，因为 w 的深度严格小于 u，不可能已作为先前路径的一部分被移除（该路径也必须经过 u，即 w唯一的直接下游节点）。

假设（11）中的事件不成立；也就是说，在 $C^{(2,\ell)}(C_{\max}(G_n(p')))$ 中有超过 $\frac{\epsilon n}{2}$ 个顶点未能加入 $C^{(2)}(C_{\max}(G_n(p)))$。我们声称，这些顶点中至少有 $\frac{3\epsilon n}{8}$ 个位于 $C^{(2,\ell)}(C_{\max}(G_n(p')))\backslash H$，因此再次被着色为黑色或紫色。为此，我们考虑两种情况：（1）如果 $H = C^{(2)}(C_{\max}(G_n(p')))$，则由 $C^{(2)}(C_{\max}(G_n(p))) \supseteq C^{(2)}(C_{\max}(G_n(p'))) = H$ 可知这一点；（2）如果$H \supsetneq C^{(2)}(C_{\max}(G_n(p')))$，则根据我们的构造，$|H| = \lfloor \frac{\epsilon n}{8} \rfloor$，且这些顶点中至少有 $\frac{\epsilon n}{2} - |H| \geq \frac{3\epsilon n}{8}$ 个位于 $C^{(2,\ell)}(C_{\max}(G_n(p')))\backslash H$，如所声称的那样。

接下来，我们声称，如果顶点 $v \in C^{(2,\ell)}(C_{\max}(G_n(p')))\backslash H$ 始终位于 $C^{(2)}(G_n(p))$ 之外，那么它的整个上游也必然如此。为此，我们通过逐一讨论以下情况来证明其逆否命题：$u \in \mathsf{us}(v) \cap C^{(2)}(G_n(p))$。

- 如果 u 位于一个环 $\Gamma \subseteq G_n(p)$ 中，那么存在一条从 $u \in \Gamma$ 到C 中某个环的（简单）路径$_{(2)}(C_{\text{最大}}(G_n(p')))$，该路径遍历了 u的整个下游。然后 v 位于环之间的一条路径上（如果不直接位于环 Γ 中）；

- 否则，u 位于两个环 Γ_1、$\Gamma_2 \subseteq G_n$之间的一条路径上（p）。因此，从 u 到 Γ_1 和 Γ_2 各有一条不相交路径，并且其中至少一条与从 u 到 v 的下游不重叠（否则，u就在一个环中）。将这条路径与 u 的整个下游拼接起来，便得到一条从 $C_{(2)}(C_{\text{最大值}}(G_n(p')))$（其中包含一个环）到 Γ_1 或 Γ_2 的简单路径，这再次证明 v位于两个环之间的一条路径上。

根据这一观察，我们可以识别出所有位于 $C^{(2)}(G_n(p))$ 之外、且下游位置最靠后的黑色和紫色顶点 $v_1,...,v_J$（即其下游不存在其他位于 $C^{(2)}(G_n(p))$ 之外的黑色或紫色顶点）；这些顶点的所有下游顶点都并入 $C^{(2)}(G_n(p))$，而其包含自身的上游部分仍在外部。特别地，任何不在 $C^{(2)}(G_n(p))$ 中的黑色或紫色顶点，都必须位于某个 $j \in [J]$ 的包含自身的上游部分中；当 (11) 中的事件不成立时，这类顶点至少有 $\frac{3\epsilon n}{8}$ 个。因此，我们必有$\sum_{j=1}^{J} |\mathsf{us}^{\mathsf{reg}}(v_j)| \geq \frac{3\epsilon n}{8}$。根据引理 G.5,$\sum_{j=1}^{J} |\mathsf{us}^{\mathsf{col}}(v_j)| \geq \frac{\epsilon n}{8}$。换言之，在 $C_{\max}(G_n(p'))$ 中存在一组彩色段，其总长度至少为$\frac{\epsilon n}{8}$，且完全位于 $C^{(2)}(G_n(p))$ 之外，正如我们所声称的。 □

命题 3.4 的证明。我们不失一般性地假设 $p > p_c$ and $\zeta^{(2,\infty)}(p) > \epsilon$。与引理 3.5 中一样，我们选择 $p' < p$，使其充分接近 p，从而 $\zeta^{(2,\infty)}(p') \geq \zeta^{(2,\infty)}(p) - \frac{\epsilon}{4}$，并选择足够大的 ℓ，使得 $\frac{1}{n}|C^{(2,\ell)}(C_{\max}(G_n(p')))| \geq \zeta^{(2,\infty)}(p') - \frac{\epsilon}{4}$and $\ell > \max\{\frac{4}{\delta}, \frac{8}{\beta^L \delta}\}$。根据引理 3.6，存在一组 S 个着色线段，其总长度至少为 $\frac{\epsilon n}{8}$，且与 $C^{(2)}(G_n(p))$ 的交集为空的概率至多为

$$2^{n/\ell} \cdot (1-\beta^L)^{\frac{\delta n}{2} - \frac{n}{\ell}} \leq e^{\frac{n}{\ell} - \beta^L \cdot \frac{\delta n}{4}} \leq e^{-\beta^L \delta n/8} \to 0 \qquad \text{as } n \to \infty. \tag{35}$$

因此，根据引理 3.5，(11) 以高概率成立，我们的证明也就完成了。 □

H 算法 1 的鲁棒性：定理 1.3 的证明

定理 1.3 的证明。按照惯例，令 μ 表示序列 $\{G_n\}_{n\in\mathbb{N}}$ 的局部极限。首先，根据定理 E.1 和 F.1，我们知道，对于所有 $p \neq p_c(\mu)$，

$$\frac{1}{n}|C^{(2)}(G_n(p))| \xrightarrow{p} \zeta^{(2)}(p;\mu) \qquad \text{and} \qquad \frac{1}{n}|C^{(2)}_{\max}(G_n(p))| \xrightarrow{p} \zeta^{(2,\infty)}(p;\mu). \tag{36}$$

因此，只要 $I^{(2)}$（分别为 $I^{(2,\infty)}$）以概率$1-\frac{\epsilon}{2}$ 位于 $\zeta^{(2)}(p)$（分别为 $\zeta^{(2,\infty)}(p)$）的 $\frac{\epsilon}{2}$-邻域内即可。目前，我们重点关注将 $I^{(2,\infty)}$ 作为$\zeta^{(2,\infty)}(p)$ 的估计；对 $I^{(2)}$ 的分析是类似的。

算法 1 在渗流图中实现了独立选取的 T 邻域，特别是，$\{I_t^{(2,\infty)}\}_{t\in[T]}$ 是独立同分布的伯努利随机变量（若输入图 G_n 是随机的，则以该输入图为条件）。对于任意 $\ell\in\mathbb{N}$（稍后指定），我们可以将 $\mathbb{E}[I_t^{(2,\infty)}|G_n(p)]=\mathbb{P}(v_t\in_{C^{(2,R_t)}}(G_n(p))|G_n(p))$ 分解为

$$\mathbb{P}(v_t\in C^{(2,R_t)}(G_n(p)),R_t\geq\ell|G_n(p))+\mathbb{P}(v_t\in C^{(2,R_t)}(G_n(p)),R_t<\ell|G_n(p)). \tag{37}$$

对于 (37) 右端的第一项，我们有

$$\mathbb{P}(v_t\in C^{(2,R_t)}(G_n(p)),R_t\geq\ell|G_n(p))\leq\mathbb{P}(v_t\in C^{(2,\ell)}(G_n(p))|G_n(p)). \tag{38}$$

由引理 B.1，$\mathbb{P}(v_t\in C^{(2,\ell)}(G_n(p))|G_n(p))\xrightarrow{p}\zeta^{(2,\ell)}(p)$，且由于 $\lim_{\ell\to\infty}\zeta^{(2,\ell)}(p)=\zeta^{(2,\infty)}(p)$，通过选择足够大的 ℓ，我们可以保证

$$\mathbb{P}\big(v_t\in C^{(2,R_t)}(G_n(p)),R_t\geq\ell\big|G_n(p)\big)\leq\zeta^{(2,\infty)}(p)+\frac{\epsilon}{8}\qquad\text{whp.} \tag{39}$$

对于第二项，注意到

$$\mathbb{P}(v_t\in C^{(2,R_t)}(G_n(p)),R_t<\ell|G_n(p))\leq\mathbb{P}(R_t<\ell|G_n(p))\leq\mathbb{P}(|\mathcal{N}_\ell(v_t;G_n(p))|>K|G_n(p)). \tag{40}$$

由引理 B.2，当 K 足够大时，该量以高概率至多为 $\epsilon/8$。因此，

$$\mathbb{E}\big[I_t^{(2,\infty)}\big|G_n(p)\big]\leq\zeta^{(2,\infty)}(p)+\frac{\epsilon}{4}\qquad\text{whp.} \tag{41}$$

另一方面，由于 $R_t\leq K$ 和 $C^{(2,R_t)}(G_n(p))\supseteq C^{(2,K)}(G_n(p))$，我们有

$$\mathbb{P}(v_t\in C^{(2,R_t)}(G_n(p))|G_n(p))\geq\mathbb{P}(v_t\in C^{(2,K)}(G_n(p))|G_n(p)). \tag{42}$$

按照与前面类似的推理，我们可以保证

$$\mathbb{E}\big[I_t^{(2,\infty)}\big|G_n(p)\big]=\mathbb{P}(v_t\in C^{(2,R_t)}(G_n(p))|G_n(p))\geq\zeta^{(2,\infty)}(p)-\frac{\epsilon}{4}\qquad\text{whp} \tag{43}$$

通过选择足够大的 K。结合 (41) 和 (43)，可得 $\big|\mathbb{E}\big[I_t^{(2,\infty)}\big|G_n(p)\big]-\zeta^{(2,\infty)}(p)\big|\leq\frac{\epsilon}{4}$ 以高概率。

其余部分是直接的。由于 $I^{(2,\infty)}$ 是 T 个独立同分布的伯努利随机变量$\{I_t^{(2,\infty)}\}_{t\in[T]}$ 的平均值（以 $G_n(p)$ 为条件），根据标准的 Hoeffding 不等式，

$$\mathbb{P}\Big(\big|I^{(2,\infty)}-\mathbb{E}\big[I_t^{(2,\infty)}\big|G_n(p)\big]\big|\geq\frac{\epsilon}{4}\Big|G_n(p)\Big)\leq\frac{\epsilon}{4} \tag{44}$$

对于 $T=O\big(\frac{1}{\epsilon^2}\log\frac{1}{\epsilon}\big)$，对 G_n 进行边际化保留了该界。因此，我们最终得到 $\big|I^{(2,\infty)}-\zeta^{(2,\infty)}(p)\big|\leq\epsilon/2$，其概率至少为 $1-\epsilon/2$。将此与 $\frac{1}{n}|C^{(2)}_{\max}(G_n(p))|$ 依概率收敛到 $\zeta^{(2,\infty)}(p)$ 结合起来，即可完成证明。

□

References

[1] David Aldous 和Russell Lyons。单模随机网络上的过程。《电子概率学报》，12：1454–1508，2007年。[2] David Aldous 和J.Michael Steele。《客观方法：概率组合优化与局部弱收敛》，第1–72页。SpringerBerlin Heidelberg，柏林，海德堡，2004年。[3] Yeganeh Alimohammadi、ChristianBorgs 和Amin Saberi。扩展图上随机有向图的局部性。arXiv预印本 arXiv:2103.09952，2021年。[4] Yeganeh Alimohammadi、ChristianBorgs 和Amin Saberi。利用局部图特征预测流行病。在《2022年ACM-SIAM离散算法年会》（SODA）论文集，第3430–3451页。SIAM，2022年。[5] Noga Alon、ItaiBenjamini和 AlanStacey。有限图上的渗流与等周不等式。《概率年刊》，32(3)：1727–1745，2004年。[6] Reid Andersen、Christian Borgs、Jennifer Chayes、John Hopcraft、Vahab S Mirrokni 和 Shang-HuaTeng。PageRank贡献值的局部计算。在网络图算法与模型研讨会（WAW），第4863卷，第150–165页。Springer，2007年。[7] Dana Angluin。处理器网络中的局部性质与全局性质。在第十二届ACM计算理论研讨会论文集，第82–93页，1980年。[8] Frank Ball、David Sirl和 Pieter Trapman。具有家庭结构的随机网络上流行病的阈值行为与最终结果。《应用概率进展》，41(3)：765–796，2009年。[9] Albert-L´aszl´o Barab´asi 和 R´eka Albert。随机网络中标度的出现。《科学》，286(5439)：509–512，1999年。[10] Itai Benjamini、AsafNachmias 和 Yuval Peres。临界渗流概率是局部的吗? 《概率论及相关领域》，149：261–269，2011年。[11] ItaiBenjamini 和 Oded Schramm。有限平面图分布极限的递归性。《OdedSchramm文选》，第533–545页，2011年。[12] B´ela Bollob´as、Svante Janson 和OliverRiordan。非齐次随机图中的相变。《随机结构与算法》，31(1)：3–122，2007年。[13] Souvik Dhara、DebankurMukherjee 和 Subhabrata Sen。配置模型中极值割的相变。《电子概率学报》，22：1–29，2017年。[14] Paul Erd˝os、Alfr´ed R´enyi等。随机图的演化。匈牙利科学院数学研究所出版物，5(1)：17–60，1960年。[15] Nikolaos Fountoulakis。具有给定度序列的稀疏随机图上的渗流。《互联网数学》，4(4)：329–356，2007年。[16] Mohsen Ghaffari。极大独立集的局部计算。在2022年IEEE第63届计算机科学基础年会（FOCS）论文集，第438–449页。IEEE，2022年。

[17] Mohsen Ghaffari、Fabian Kuhn 和 Yannic Maus。《局部分布式图问题的复杂性》。载于《第49届 ACM SIGACT 计算理论年会论文集》，第784–797页，2017年。[18]Transforma Insights。2019年至2021年全球物联网（IoT）联网设备数量，以及2022年至2030年的预测（单位：十亿）。可访问：https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/，2022年7月。[19] Svante Janson 和 Malwina J Luczak。 k-核问题的一个简单解。《随机结构与算法》，30(1-2):50–62，2007年。[20] Michael Krivelevich、Eyal Lubetzky 和 Benny Sudakov。高围长扩展图上的渗流渐近性。《随机结构与算法》，56(4):927–947，2020年。[21] Fabian Kuhn、Thomas Moscibroda 和 Roger Wattenhofer。《近视的代价》。载于 SODA 2006：第17届 ACM-SIAM 离散算法研讨会，美国佛罗里达州迈阿密，2006年。[22] Fabian Kuhn、Thomas Moscibroda 和 Rogert Wattenhofer。《什么无法在局部计算！》。载于《第23届 ACM 分布式计算原理年会论文集》，第300–309页，2004年。[23] Reut Levi、Dana Ron 和 Ronitt Rubinfeld。《稀疏生成图的局部算法》。Algorithmica，82(4):747–786，2020年。[24] Don R Lick 和 Arthur T White。 k-退化图。《加拿大数学杂志》，22(5):1082–1096，1970年。[25] Nathan Linial。《分布式图算法中的局部性》。SIAM 计算杂志，21(1):193–201，1992年。[26] Peter Macgregor 和 He Sun。《寻找密集连接簇的局部算法》。载于《国际机器学习会议》，第7268–7278页。PMLR，2021年。[27] Yishay Mansour 和 Shai Vardi。《最大匹配的局部计算近似方案》。载于《第16届组合优化问题近似算法国际研讨会（APPROX）2013暨第17届随机化与计算国际研讨会（RANDOM）2013》，美国加利福尼亚州伯克利，2013年8月21日至23日。论文集，第260–273页。Springer，2013年。[28] Sharon Marko 和 Dana Ron。《有界度图和一般稀疏图中的距离近似》。载于《第9届组合优化问题近似算法国际研讨会（APPROX）2006暨第10届随机化与计算国际研讨会（RANDOM）2006》，西班牙巴塞罗那，2006年8月28日至30日。论文集，第475–486页。Springer，2006年。[29]Meta。2022年第四季度财报演示。可访问：https://s21.q4cdn.com/399680738/files/doc_financials/2022/q4/Earnings-Presenta

tion-Q4-2022年

2023年1月。[30] Moni Naor 和 Larry Stockmeyer。局部计算可以实现什么？载于《第二十五届 ACM计算理论研讨会论文集》，第 184–193 页，1993 年。

[31]Huy N Nguyen 和 Krzysztof Onak。通过局部改进实现常数时间近似算法。载于《2008 年第 49 届 IEEE计算机科学基础研讨会论文集》，第 327–336 页。IEEE，2008 年。[32] Michal Parnas 和 Dana Ron。次线性时间近似最小顶点覆盖及其与分布式算法的联系。《理论计算机科学》，381(1-3):183–196，2007年。[33] Merav Parter、Ronitt Rubinfeld、Ali Vakilian 和 Anak Yodpinyanee。伸展子图的局部计算算法。arXiv 预印本 arXiv:1902.08266，2019年。[34] Boris Pittel、Joel Spencer 和 Nicholas Wormald。随机图中巨型 k-核的突然出现。《组合理论杂志 B辑》，67(1):111–151，1996年。[35] Derek J de Solla Price。科学论文网络：参考文献模式揭示科学研究前沿的性质。《科学》，149(3683):510–515，1965年。[36] Filippo Radicchi、Claudio Castellano、Federico Cecconi、Vittorio Loreto 和 Domenico Parisi。网络中社群的定义与识别。《美国国家科学院院刊》，101(9):2658–2663，2004年。[37]Oliver Riordan。 k-核与分枝过程。《组合学、概率与计算》，17(1):111–136，2008年。[38] Ronitt Rubinfeld、Gil Tamir、Shai Vardi 和 Ning Xie。快速局部计算算法。arXiv 预印本 arXiv:1104.1377，2011年。[39] Sourav Sarkar。关于一列扩展图的局部弱极限的注记。《电子概率通讯》，26:1–6，2021年。[40] Johannes Schneider 和 Roger Wattenhofer。适用于增长有界图的对数星分布式极大独立集算法。载于《第二十七届 ACM分布式计算原理研讨会论文集》，第 35–44 页，2008年。[41] Stephen B Seidman。网络结构与最小度。《社会网络》，5(3):269–287，1983年。[42] Pieter Trapman。网络上流行病的分析方法。《理论种群生物学》，71(2):160–173，2007年。[43] J. van den Berg 和 Michael Keane。关于渗流概率函数的连续性。《当代数学》，26:61–65，1984年。[44]Remco van der Hofstad。随机图中的巨型分量几乎是局部的。arXiv 预印本 arXiv:2103.11733，2021年。[45] Remco van der Hofstad、Johan SH van Leeuwaarden 和 Clara Stegehuis。层次配置模型。《互联网数学》，第 1214 页，2016年。