随机单纯复形的强塌缩

Jean-Dan iel Boissonnat [*1]、Kunal Dutta [†2]、Soumik Dutta [†3]和 Siddharth Pritam [*4]

[1]法国索菲亚·安提波利斯，蔚蓝海岸大学、INRIA。邮箱：Jean-Daniel.Boissonnat@inria.fr[2]波兰华沙大学数学、信息学与力学学院。邮箱：K.dutta@mimuw.edu.pl[3]波兰华沙大学数学、信息学与力学学院。邮箱：s.dutta2@mimuw.edu.pl[4]法国索菲亚·安提波利斯，蔚蓝海岸大学、INRIA。邮箱：siddharth.pritam@inria.fr

arXiv:2301.03514v1 [cs.CG] 9 Jan 2023

摘要

Barmak 和 Minian [6], 提出的单纯复形的强塌缩，是将一个复形组合塌缩到其子复形上的过程。近年来，计算拓扑学研究者 [22, 7, 8], 对其给予了关注，因为实证观察发现，在保持同伦类的同时，它有助于简化单纯复形并缩减其规模。我们研究随机单纯复形上的强塌缩过程。对于具有 n 个顶点、边概率为 c/n （其中 $c > 1$）的 Erd˝os-R´enyi 随机团复形 $X(n, c/n)$，我们证明，经过任意一个极大强塌缩序列后，剩余的子复形（即核）渐近几乎必然具有 $(1-\gamma)(1-c\gamma)n + o(n)$ 个顶点，其中 γ 是函数 $f(x) = \exp(-c(1-x))$ 在区间 $(0, 1)$ 内最小的非负不动点。这是关于随机（或非随机）单纯复形上强塌缩的首批理论结果。

1 引言

研究动机简单坍缩是一种组合概念，可以在不改变单纯复形拓扑结构的情况下对其进行简化。它可以表示为一系列基本移动，每次移除一对单纯形 σ和 τ，使得 σ唯一地包含于 τ中。J.H.C. Whitehead [21] 提出了简单坍缩的概念，用于研究胞腔复形的同伦类型。此后，这一概念在拓扑学的许多不同领域中得到应用，尤其是在计算拓扑学中。近年来，人们提出了简单坍缩的新变体，称为强塌缩，更一般地称为 d-塌缩 [6, 8, 5]。在这类塌缩中，人们会移除一些特殊顶点（更一般地说，是d-单纯形），称为被支配顶点（单纯形），它们的链接是一个单纯锥。这同样可以表示为一系列基本移动，每次移除被支配顶点（单纯形）。事实证明，这些方法是解决计算拓扑学中许多问题的强大工具。尤其是，Pritam 等人近期的研究 [9, 8, 23] 表明，强塌缩和边塌缩（对于 $d = 1$而言，d-塌缩）可用于高效计算单参数及

[*]J.-D. Boissonnat 和 S. Pritam 获得了欧洲研究委员会（ERC）通过欧盟第七框架计划（FP/2007- 2013）提供的资助，资助协议编号为 339025 GUDHI（高维几何理解的算法基础）。

[†]K. Dutta 和 S. Dutta 获得了波兰国家科学中心 SONATA 项目资助，项目编号为 2019/35/D/ST6/04525。

多参数持久性。高效计算持久同调是拓扑数据分析领域的核心研究课题之一。

持久同调的计算涉及计算一列嵌套单纯复形的同调群，这列单纯复形称为滤过。计算持久同调需要$\mathbb{O}(n^3)$时间和 $\mathbb{O}(n^2)$空间，其中 n 是滤过中的单纯形总数。 [9, 7, 8, 12, 23] 中提出的通用技术是使用强塌缩或边塌缩将滤过化简为更小的滤过，同时保留持久同调。
[9, 7, 8, 12, 23],通过实验表明，在实际应用中，化简后的滤过非常小，因此持久同调的计算极快。对于旗（团）复形，效率提升尤为显著，因为只需利用给定复形 [7, 12, 23]的图（1-骨架），即可计算强塌缩和边塌缩。

如上所述， [9, 7, 12, 23] 中报告的效率是通过实验得出的，而化简规模并无理论保证。这是因为通常化简程度取决于具体的复形及其组合结构。事实上，化简甚至取决于塌缩的顺序；不同的顺序可能得到不同的核，强塌缩的情况除外。对于滤过单纯复形，研究化简规模则更加困难。这促使我们考虑随机单纯复形，并研究塌缩所带来的平均化简规模。

本文研究定义在 Erd˝os-R´enyi 随机图上的团复形经过强坍缩后所达到的约化规模问题。

前人 Linial 和 Meshulam 在其开创性论文 [16] 中开启了随机单纯复形的研究。随后，Meshulam 和 Wallach [19] 将随机复形模型推广，提出了 d维随机复形的 $Linial$-$Meshulam$ (LM) 模型。此后，众多作者围绕许多不同的随机单纯复形模型开展了大量研究，探讨这类复形的各种拓扑和几何性质 [14]。随机单纯复形的简单坍缩也引起了研究者的兴趣，并产生了大量相关工作。在 d维 LM 模型中，Kozlov [15] 证明了第 d阶同调消失阈值的界。Aronshtam、Linial、 Łuczak 和 Meshulam [4], 最先研究随机复形上的简单坍缩，他们改进了 Kozlov 的界，得到了紧确的阈值界，并给出了 d维 LM 模型中可坍缩性的阈值界。随后，Aronshtam 和 Linial [2, 3] 进一步推进了这项研究，先在 [2] 中得到了第 d阶同调消失的阈值，继而得到了 d维 LM 复形 [3]不可坍缩的阈值。在 [17] 中，Linial 和 Peled 得到了这类复形的核大小的精确渐近界。最近，Malen [18] 证明，当 $\alpha > 1/(k+1)$ 时，ER 团复形 $X(n,p)$ 以高概率是 $(k+1)$-可坍缩的。

因此，据我们所知，迄今为止关于随机复形中坍缩的研究仅涉及简单坍缩。

本文中，我们将用渐近几乎必然（a.a.s.）表示一系列事件$(E_n)_{n\geq 1}$，当 E_n 发生的概率随着 $n \to \infty$ 趋于 1 时。

随机单纯复形模型 本文将考虑以下两种随机单纯复形模型。对于一个图 G，令 cl(G) 表示 G 上的团（旗）复形，即这样一种单纯复形： G 上每个由 d 个顶点构成的完全子图都是 cl(G) 中的一个 $d-1$-单形。具有概率参数 $p \in [0,1]$ 、定义在 n 个顶点上的$Erd\tilde{o}s$-$R\acute{e}nyi$ （ER）模型 $X(n,\ p)$，通过以概率 p 随机且独立地将一个含 n 个元素的集合中的每一对可能元素连接成一条边，得到随机图 $G = G(n,\ p)$。令 $X(n,\ p) := \mathrm{cl}(G(n,\ p))$ 。

我们的贡献 我们给出了随机单纯复形经过强（顶点）塌缩后其核 （即不含任何被支配顶点的最小子复形）大小的界限。以往的研究着重于计算不同 k 下 k 阶同调类出现和消失的阈值，而本文更关注核的大小。我们证明，对于具有 n 个顶点的 ER 团复形，经过一系列极大强塌缩后，其核的大小几乎必然是 n 的一个常数比例；该常数仅取决于边概率，且与 1 保持正距离。此外，我们还给出了该常数的精确表达式，它是一个隐式方程的不动点。下面陈述我们的第一个定理。

设 K 是一个单纯复形。对于 $i \geq 1$，我们用 $f_i(K)$ 表示该复形的 $i-$ 单纯形集合。稍微滥用一下记号，我们令 $f_0(K)$ 表示该复形的非孤立顶点集合~，并令 $f_0(K)$ 表示该复形的非孤立顶点集合。（若顶点$v \in K$ 不是 K 中任何一条边的面，则称其为孤立顶点。）在对 K 执行的剪枝阶段中， K 中所有被支配的（强可坍缩）顶点会同时坍缩。令 $R_t(K)$ 表示对 K 执行 t 个剪枝阶段后得到的复形。最后， $R_\infty(K)$ 表示对 K 执行 K 个强塌缩的极大序列后得到的复形，即 K 的核。

定理 1. 设 $c > 1$ 和 $X \sim X(n, c/n)$。那么存在常数 $\gamma \equiv \gamma(c)$，它由函数 $f(x) = \exp(-c(1-x))$ 的最小非负不动点给出，且 $x \in (0,1)$，使得 $0 < \gamma < 1/c < 1$ 并且以下结论以渐近概率 *1* 成立

$$|f_0(R_\infty(X))| = (1-\gamma)(1-c\gamma)n + o(n).$$

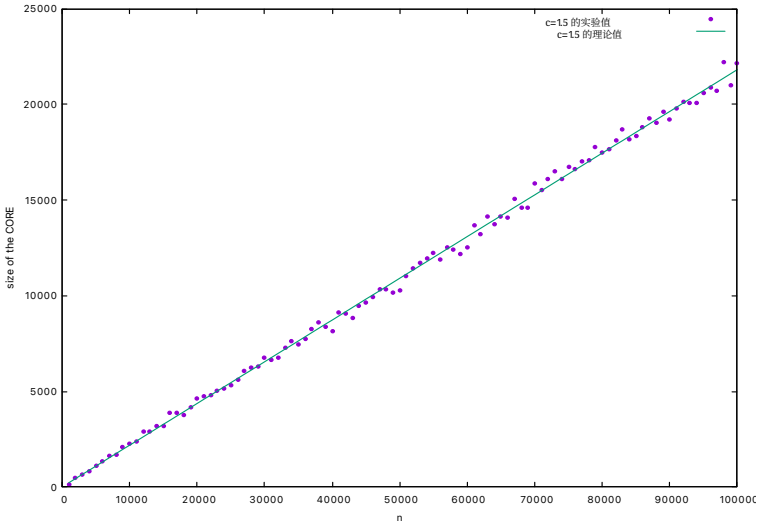


图 1：我们对 ER复形进行了实验，并绘制了核的大小（强塌缩），其中使用了 $n \in [0, 10^5]$ 和 $c = 1.5$。这些实验清楚地验证了我们的理论结果。

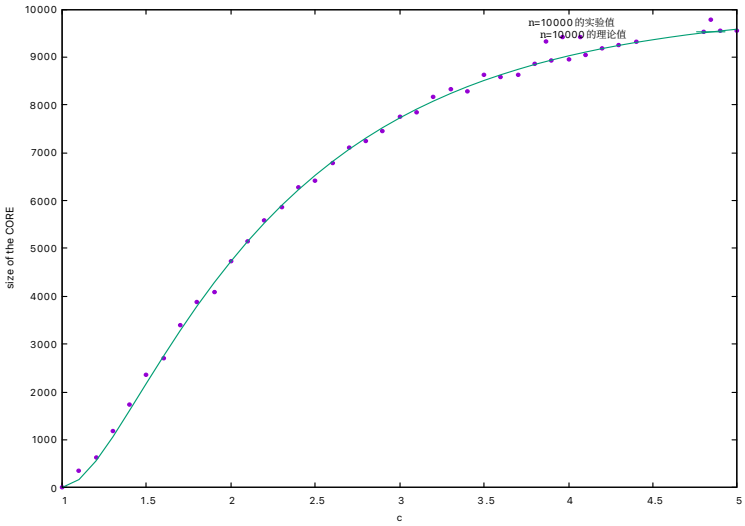


图2：在一组不同的 ER复形实验中，我们改变了常数 $c \in [1,5]$，并将顶点数固定为 $n = 10^4$。

接下来我们讨论一个与算法相关的问题：给定 $\varepsilon \in (0,1)$，在第一阶段中需要多少轮次才能使结果与核的实际大小相差不超过 εn ？以下定理给出了轮次数量 t 关于 ε的函数界。

定理 2. 设 $X \sim X(n, c/n)$ 和 $\varepsilon > 0$ （充分小）给定。则存在$t \in \mathbb{Z}_+$ 使得

$$\frac{(c-c\gamma)e^c}{1-ce^{-c}}(ce^{-c})^t \le \epsilon \le \frac{(c+1)e^c}{1-c\gamma}(c\gamma)^t,$$

且 $|f_0(R_t(X))| - |f_0(R_\infty(X))| \le \epsilon n + o(n)$ 几乎必然成立。

1.1 证明概述与各节安排

虽然我们的分析总体上沿用了例如 [4, 2, 3]，中对 LM复形简单可塌缩性的分析流程，但仍存在若干差异和难点。首先，在这两种情形下（ER 团复形的强塌缩和 LM模型的简单坍缩），我们的目标都是确定核的大小，而非判断复形是否可坍缩。其次，对于随机 ER 团复形，我们的分析需要考虑复形的非齐次性。也就是说，该模型中的极大单形大小可能各不相同。此外，与LM模型不同，一个极大单形的存在与其他所有可能的极大单形是否存在并不相互独立。最后，在我们的研究背景下，随机 ER 团复形模型最有意思的差异或许是，移除一个顶点所产生的影响未必局限于局部——这一事实需要我们在多个地方进行相当程度的创新来加以处理，尤其是在第 6 节证明集中界时，以及分析的后续阶段（第7节）。我们在第 6 节开头对用于证明集中界的证明策略作了更详细的概述。

谨记上述注意事项，我们先简要回顾 Aronshtam 和 Linial [3]证明的主要思路。分析分为两个阶段，每个阶段又进一步划分为若干轮次（阶段）。在第一阶段的每一轮中，所有简单可塌缩单形都同时坍缩，并将这一过程重复固定轮数。Aronshtam 和 Linial [3] 使用随机单纯复形的树状模型来近似随机LM复形的局部结构，并证明，随着轮次数增加，所有这些轮次中塌缩单形的总数趋于初始单纯形数量的一个固定比例

。此外，这个极限常数可以表示为一个隐式方程的不动点，该方程仅涉及分布参数 c。

在第二阶段的分析中，每一轮都会从（互不相邻的）简单可折叠单形集合中随机选取一个单形并将其坍缩。我们的目标是证明，在这一阶段中，坍缩的单形数量相对于最初存在的单形数量而言渐近可忽略。因此，概括来说，最终删除的单形数量由第一阶段本身决定，而第二阶段则用于证明这一界是紧的。

我们的证明也将分析分为两个阶段。与 [3], 类似，我们证明，随机单纯复形的一种树状模型能够很好地近似局部邻域；这部分内容见第 3 节。接下来是第 4 节对第一阶段的分析。本节的主定理给出了第一阶段经过 t 轮（或剪枝阶段）后剩余顶点数量的期望表达式。关于 t 在 ε、γ 和c条件下的界，见定理 2；该定理在第 5 节中证明。不过，在开始分析第二阶段之前，我们还需要得到核的大小以及若干其他随机变量的集中界。这些结论在第 6 节中证明，其中我们使用临界和预临界（子）复形 的概念——第 6 节开头对此有更详细的介绍。得到这些集中界后，我们便在第7节分析第二阶段。

2 预备知识

本节简要介绍一些拓扑学和概率概念。关于拓扑学相关主题的全面介绍，读者可参阅 [13] ；关于概率论和随机结构相关主题，可参阅 [10] 。

单纯复形。 一个**抽象单纯复形** K 是非空有限集X, 的子集构成的集合，并且满足：对于K中的每个子集A ， A 的所有子集都属于K。K 中的元素称为一个**单形**。基数为 $d+1$ 的元素称为一个 d-单形，而 d称为其**维数**。给定一个单纯复形 K，我们将其几何实现记作$|K|$。如果一个单形不是K中任何其他单形的真子集，则称其为**极大**单形。若K 的子集L 本身也是一个单纯复形，则称其为一个**子复形**。如果 的子复形K' 包含 中维数至多为 d的所有单形，则称其为**骨架** 。

Erdős–Rényi 图定义。 这是由所有定义在 n 个顶点上的图组成的概率空间 $G(n,\ p)$。具有 m 条边的图出现的概率为 $p^m(1-p)^{(n)(n-1)/2-m}$。换言之，这是一个定义在 n 个顶点上的随机图，其中每条边都以 p 的概率独立出现。

团复形与邻域。若复形 K 中某个顶点子集构成一个团（即任意两个顶点之间都有一条边相连），且这些顶点张成一个单形，则称该复形为团复形或旗复形。对于 G 中的顶点 v ， G 中 v 的开邻域 $N_G(v)$ 定义为 $N_G(v) := \{u \in G \mid [uv] \in E\}$，其中 E 是 G 的边集。闭邻域 $N_G[v]$ 为 $N_G[v] := N_G(v) \cup \{v\}$。类似地，我们将边 $[xy] \in G$ 的闭邻域和开邻域分别定义为 $N_G[xy]$ 和 $N_G(xy)$，即 $N_G[xy] := N[x] \cap N[y]$ 和 $N_G(xy) := N(x) \cap N(y)$。上述定义可以推广到 G 的任意 k-团 $\sigma = [v_1, v_2, ..., v_k]$ ； $N_G[\sigma] := \bigcap_{v_i \in \sigma} N[v_i]$ 和 $N_G(\sigma) := \bigcap_{v_i \in \sigma} N(v_i)$。

星、链接与单纯锥。 设 σ 是单纯复形 K中的一个单形，则 σ在 K中的**闭星** $st_K(\sigma)$是 K的一个子复形，定义如下： $st_K(\sigma) := \{\tau \in$

$K|\ \tau\cup\sigma\in K\}$. σ在 K中的**链接** $lk_K(\sigma)$定义为 $st_K(\sigma)$中与 σ不相交的单形集合： $lk_K(\sigma):=\{\tau\in st_K(\sigma)$ $|\tau\cap\sigma=\emptyset\}$. σ在 K中的**开星** $st^o_K(\sigma)$定义为集合 $st_K(\sigma)\setminus lk_K(\sigma)$。通常，$st^o_K(\sigma)$不是 K的子复形。

设 L 是一个单纯复形， a 是一个不属于 L的顶点。则集合 aL 定义为$aL:=\{a,\tau\mid\tau\in L\ or\ \tau=\sigma\cup a;$ 其中 $\sigma\in L\}$ 称为**单纯锥**。

简单坍缩。 给定一个复形 K，若单形 $\sigma\in K$ 满足 σ 有唯一的余面 $\tau\in K$，则称其为**自由单纯形** 。对 $\{\sigma,\tau\}$ 称为**自由对**。移除自由对的操作： $K\to K\setminus\{\sigma,\tau\}$ 称为**初等简单坍缩**。一系列这样的初等简单坍缩称为**简单坍缩**。记作 $K\searrow L$。若子复形 K^{ec} 是 K 的**初等核**，则 $K\searrow K^{ec}$ ，且 K^{ec} 没有自由对。

单形的移除。 我们用 $K\setminus\sigma$ 表示通过移除σ从 K 得到的子复形，即包含 K 的所有单形（单形 σ 及 σ的余面除外）的复形。

受支配单纯形。 如果单形 σ in K 的链接 $lk_K(\sigma)$ 是单纯锥，则称 σ 为**受支配单纯形** ，也就是说，存在一个顶点 $v'\notin\sigma$ 和 K 的子复形 L ，使得 $lk_K(\sigma)=v'L$。我们称顶点 v' 支配 σ ，并称 σ 被 v'支配 ，记作$\sigma\prec v'$。

σ-代数 读者可参阅 [10] ，了解 σ-代数的定义。

d-坍缩。 给定一个复形 K，从 K 中移除一个受支配的 k-单形 σ ，称为一次**初等 d-坍缩**，记作 $K\searrow\!\searrow^d\{K\setminus\sigma\}$。一系列初等 d-坍缩称为一次 **d-坍缩**，记作 $K\searrow\!\searrow^k L$。如果一个复形 K d没有任何受支配的 d 单纯形，则称其为**-坍缩极小**。若 $K\searrow\!\searrow^k K^d$ 且 K^d 是 d-坍缩极小，则称 K 的子复形 K^k 为一个 **d-核** 。复形 K 的 0-核是唯一的，但对于 $k\geq1$，通常并非如此。与简单坍缩一样， d-坍缩保持单纯复形的同伦类型。

0-塌缩是 [6] 中引入的**强塌缩** ，而 1-塌缩称为**边塌缩** [8]。 [8] 中的以下引理用邻域刻画了旗复形这一特殊情形下单形的支配关系。

引理 3. 设 σ 是旗复形 K 的一个单形。那么，当且仅当 $N_G[\sigma]\subseteq N_G[v']$ 时，顶点 v' 支配 σ 。

本文主要关注 $d=0$ 的情形，即 σ 是一个顶点。 [7], 中的下一个引理虽然初等，却至关重要。

引理 4. 设 K 是一个旗复形， L 是 K 经强塌缩得到的任意子复形。那么 L 也是一个旗复形。

这两个引理（引理 3 和引理 4）表明，强塌缩非常适用于旗复形。在接下来的章节中，我们将研究 Erd˝os-R´enyi 随机图的团复形的强塌缩能力。

3 树过程

本节将介绍树过程，该过程用于模拟（强）坍缩第一阶段中的塌缩过程。一维树按如下方式递归构建:

1. 从一个节点（根）开始。

2. 在第 n 次迭代中，从参数为 c（>1）的泊松分布中为所有距离根 $n-1$ 的叶节点添加子节点。

令 $\mathcal{T}_n$ 表示经过 n 次迭代后，$n>1$ 和 $\mathcal{T}_0$ 本身为根时所有可能的树的集合。令 $\mathcal{T}:=\bigcup \mathcal{T}_n.n\in\mathbb{N}$

令 γ_t 表示树 $T\in\mathcal{T}_t$ 经至多 $t-1$ 步剪枝后缩减为根的概率。显然，$\gamma_1=e^{-c}$。设 $\gamma_0=0$。此外，我们还有如下普遍成立的递归关系：

$$\gamma_{t+1}=e^{-c(1-\gamma_t)}.$$

请注意，在这一过程中，即使根的度为 1，我们也绝不会剪去根本身。我们称这样的过程为根折叠。令 γ_t^k 表示树 $T\in\mathcal{T}_t$ 经过 $t-1$ 步根折叠后度为 k 的概率。那么，

$$\gamma_t^k=\frac{(c(1-\gamma_{t-1}))^k}{k!}e^{-c(1-\gamma_{t-1})}$$

注意，γ_1^k 给出了初始度分布。此外，令 $\gamma_t^{\geq 2}$ 表示一个顶点经过 $t-1$ 步根折叠后度至少为 2 的概率。那么我们有

$$\gamma_t^{\geq 2}=\sum_{k=2}^{\infty}\gamma_t^k=1-\gamma_t(1+c(1-\gamma_{t-1})).$$

定义 $\beta_t:=1-\gamma_{t+1}$。因此，β_t 是树 $T\in\mathcal{T}_t$ 至少需要 $t+1$ 步根折叠才能隔离根节点的概率。

定义 $f(x):=e^{-c(1-x)}$ 在区间 $[0,1]$上。除非另有说明，本文其余部分均假设 $c>1$。注意，$f([0,1])\subseteq[0,1]$、$f'(x)=cf(x)$，且 $f(x)$ 在区间 $[0,1]$上严格递增。令 $f_t(x)$ 表示由复合 f t 次得到的函数。于是，对于所有 $t\geq 1$，f_t 在 $[0,1]$ 上也严格递增。由于 $\gamma_1>\gamma_0$，在等式两边应用 f_{t-1}，可得对于所有 $t\geq 1$，$\gamma_t>\gamma_{t-1}$。

另定义 γ 为函数 $g(x):=e^{-c(1-x)}-x$ 在范围 $[0,\infty)$ 上的最左侧零点。注意，当 $g(1)=0$ 时，$0<\gamma\leq 1$。

引理 5. 对于如前所定义的 $c>1$ 和 $\gamma=\gamma(c)$，有 $c\gamma<1$。

引理 5 的证明。设 $g(x)$ 如上所定义。因此，$g(\gamma-)>0$，且 $g(\gamma)=0$。于是，由于 $g(x)$ 的可微性，$c\gamma-1=g'(\gamma)\leq 0$。现在，若 $c\gamma-1=0$，则 $ce^{1-c}=1$，这对于 $c>1$ 不可能成立。因此 $c\gamma-1<0$。 □

现在注意到，$x\leq\gamma\implies f(x)\leq f(\gamma)=\gamma$。因此，由于 $f'(x)=cf(x)$，限制在 $[0,\gamma]$ 上的 $f(x)$ 是一个压缩映射。因此，根据巴拿赫不动点定理，$f|_{[0,\gamma]}$ 有唯一的不动点，在我们的情形中，该点为 γ。

总结以上论证，我们得到如下注释。

注释 1. γ_t 作为递增序列收敛到 $0<\gamma<1/c$，而 β_t 作为递减序列收敛到$1-1/c<\beta<1$。

4 第一阶段

本节将分析坍缩的第一阶段。第一阶段分多个阶段执行，在每个阶段中，我们同时移除一个被支配顶点的极大集合。

本节的目标是证明以下定理。

定理 6. 设 $X \sim X(n, c/n)$。令 $\mathbb{E}(|f_0(R_t(X))|)$ 表示经过 X 次强塌缩阶段后 t 非孤立顶点的期望数量，并设 γ_t 如上一节所定义。那么，

$$\mathbb{E}(|f_0(R_t(X))|) = (1 - \gamma_{t+1} - c\gamma_t + c\gamma_t^2)n.$$

我们首先证明一些关于复形局部结构的重要引理。对于$X \sim X(n, \frac{c}{n})$，定义如下事件，

$$D := \{deg(v) \leq \log n \ \ \forall v \in \tilde{f_0}(X)\}.$$

然后，可以使用标准的切尔诺夫界证明以下引理。

引理 7。 $Pr\{D\} = 1 - o_n(1)$。

引理 7 的证明。注意，对于任意 $v \in f_0(X)$， $deg(v) \sim Bin(n-1, c/n)$。令 D_v 表示事件 $deg(v) \leq \log(n)$。那么根据二项分布的切尔诺夫界

$$\begin{aligned} Pr\{\neg D_v\} &= Pr\{deg(v) > \log(n)\} \\ &\leq (\frac{ec(n-1)}{n\log(n)})^{\log(n)} \\ &\leq (\frac{ec}{\log(n)})^{\log(n)} \\ &= n^{-\log(\log(n/ec))} \end{aligned}$$

因此，根据并集界 $Pr\{\bigcup_{v\in f_0(X)} \neg D_v\} \leq n^{1-\log(\log(n/ec))}$。于是，

$$Pr\{D\} = Pr\{\bigcap_{v\in f_0(X)} D_v\} = 1 - Pr\{\bigcup_{v\in f_0(X)} \neg D_v\} \geq 1 - n^{1-\log(\log(n/ec))} = 1 - o_n(1)$$

□

我们用 $Cl(S)$ 表示集合 S 的单纯闭包。固定 $v \in f_0(X)$。定义 $\mathcal{N}_0 := Cl(v)$ 和 $\mathcal{N}_{-1} := \emptyset$。另外定义 $\mathcal{N}_{i+1} := Cl(\{s \in X | \exists u \in f_0(\mathcal{N}_i) \setminus f_0(\mathcal{N}_{i-1}) | u \subset s\}) \cup \mathcal{N}_i$。等价地，也可以根据复形的 1-骨架来定义这一点。

定义事件 $\mathcal{A}_t = \{\mathcal{N}_t \in \mathcal{T}\}$。

引理 8。 令 $X \sim X(n,p)$ 并固定 $v \in f_0(X)$。那么 $Pr\{\mathcal{A}_t \cap D\} = 1 - o(1)$。

证明。 如果 $X \in D$，那么 $\tilde{f_0}(\mathcal{N}_t) = O(log^{t+1}n)$，并且我们希望避免 $O(log^{2t+2}n)$ 条边，从而使 $\mathcal{N}_t$成为一棵一维树。发生这种情况的概率为

$$(1 - c/n)^{O(log^{2t+2}n)} = 1 - o(1)$$

□

现在，这棵树中一个节点的度来自 $Bin(n-1, c/n)$。当 n 较大时，这个分布可以用 $Po(c)$ 近似。

定理 6 的证明。回忆一下，对于一个单纯复形 X， $R_t(X)$表示经过 t 个阶段后得到的复形，而 $f_0(X)$ 表示该复形的非孤立顶点（0–个单纯形）集合。注意，如果一个顶点 $v \in f_0$（X）在经过 t 个剪枝步骤后仍然存活，那么在经过 deg（v）个剪枝步骤后，它的度必定为 ≥ 2（$t-1$）。因此，$Pr\{v \in f_0$（R_t（X））$\} \leq \gamma_t^{\geq 2} = 1 - \gamma_t$（$1 + c$（$1 - \gamma_{t-1}$））。这个事件同时统计了孤立顶点和度为一的顶点（即可坍缩顶点），因此会略微高估。为了得到更精确的估计，我们注意到，按照 [3], 的思路，若一个顶点 $v \in f_0$（X）在经过 t 个剪枝步骤后既未坍缩也未成为孤立顶点，那么它就能存活 t 个剪枝步骤。此类事件发生的概率为 $1 - \gamma_{t+1} + c\gamma_t - c\gamma_t^2$。前一个引理断言，这确实是单纯复形中一个顶点的存活概率。 □

5 收敛速率

在本节中，我们证明定理 2，从而给出变量γ_t的收敛速率界。

引理 9. 设 c、γ_t 如前所定义。则

$$c\gamma_t \le \frac{\gamma_{t+1}-\gamma_t}{\gamma_t-\gamma_{t-1}} \le c\gamma_{t+1}$$

证明。设 $f(x)$ 和 $g(x)$ 如第 3 节所定义。显然,

$$\begin{aligned}
\frac{\gamma_{t+1}-\gamma_t}{\gamma_t-\gamma_{t-1}} &= \frac{g(\gamma_t)}{g(\gamma_{t-1})} \\
&= \frac{g(\gamma_{t-1}+(\gamma_t-\gamma_{t-1}))}{g(\gamma_{t-1})} \\
&= \frac{g(\gamma_{t-1})+g'(s)(\gamma_t-\gamma_{t-1})}{g(\gamma_{t-1})} && (\text{对某个 } s\in[\gamma_{t-1},\gamma_t]) \\
&= f'(s) = cf(s) && (\text{因为 } \gamma_t-\gamma_{t-1}=\gamma_{t-1})
\end{aligned}$$

由于 $f(x)$ 是递增函数，结论得证。 □

特别地，对于所有 $t\ge$ ~~1, $ce^{-c} \le \frac{\gamma_{t+1}-\gamma_t}{\gamma_t-\gamma_{t-1}} \le c\gamma$~~。

设 $\delta(t) = (1-\gamma_{t+1}-c\gamma_t+c\gamma_t^2) - (1-\gamma_{t+2}-c\gamma_{t+1}+c\gamma_{t+1}^2)$，如第7节中所定义。可以证明

$$\begin{aligned}
\delta(t) &= (1-\gamma_{t+1}-c\gamma_t+c\gamma_t^2) - (1-\gamma_{t+2}-c\gamma_{t+1}+c\gamma_{t+1}^2) \\
&= (\gamma_{t+2}-\gamma_{t+1}) + c(\gamma_{t+1}-\gamma_t) + c(\gamma_t+\gamma_{t+1})(\gamma_t-\gamma_{t+1}) \\
&= (\frac{\gamma_{t+2}-\gamma_{t+1}}{\gamma_{t+1}-\gamma_t} + c - c(\gamma_{t+1}+\gamma_t))(\gamma_{t+1}-\gamma_t)
\end{aligned}$$

因此,

$$(c-c\gamma)e^c(ce^{-c})^t \le \delta(t) \le (c+1)e^c(c\gamma)^t$$

现在定义 $\epsilon \equiv \epsilon(t) := (1-\gamma_{t+1}-c\gamma_t+c\gamma_t^2) - (1-\gamma-c\gamma+c\gamma^2)$，使得 $\mathbb{E}[|f_0(R_t(X))|] - (1-\gamma)(1-c\gamma] = \epsilon n$。因此,

$$\begin{aligned}
\epsilon(t) &= (1-\gamma_{t+1}-c\gamma_t+c\gamma_t^2) - (1-\gamma-c\gamma+c\gamma^2) \\
&= (\gamma-\gamma_{t+1}) + c(\gamma-\gamma_t) + c(\gamma_t+\gamma)(\gamma_t-\gamma) \\
&= (\gamma-\gamma_{t+1}) + (c-c(\gamma_t+\gamma))(\gamma-\gamma_t)
\end{aligned}$$

So,

$$\frac{(c-c\gamma)e^c}{1-ce^{-c}}(ce^{-c})^t \le \epsilon(t) \le \frac{(c+1)e^c}{1-c\gamma}(c\gamma)^t.$$

由此得到以下推论。

推论 10。对于任意 $t\ge 1$

$$e^c(ce^{-c})^t \le \gamma_{t+1}-\gamma_t \le e^c(c\gamma)^t$$

and

$$\frac{e^c}{1-ce^{-c}}(ce^{-c})^t \le \gamma-\gamma_t \le \frac{e^c}{1-c\gamma}(c\gamma)^t.$$

由上述推论，我们得到定理 2。

6 Co第一阶段后复形大小的集中性

本节将证明核大小的一个集中界。与 d维 LM复形中的简单坍缩不同， [2, 3], 在我们的情形下，集中界不那么容易得到。首先注意，删除单个顶点 v 可能会改变任意多个顶点的支配状态，例如，当 v 支配整个复形时便是如此。因此，顶点的影响在最坏情况下可能是 n 。因此，我们需要使用边暴露鞅不等式；该不等式是 Freedman [11],提出的一条不等式的变体，由 Warnke [20],给出，它允许我们考察单条边所产生影响的路径方差 ，而非其最坏情况影响。

为了界定路径方差，我们将证明：如果某个变量的影响很大，那么必定存在一类特定的子复形，我们称之为临界复形，其中至少有一个会出现在该复形的 1-骨架中。不难证明（我们也会给出证明），这些子图在原始随机复形中出现的概率极低。然而，在边暴露鞅的每一步都需要控制方差，也就是说，即使是在计算原始复形的任意小子复形上的期望时，也需要控制方差。为此，我们需要定义临界复形的一个超集，称为前临界复形，并证明在边暴露过程中，它们出现的概率始终极低。接下来，我们可以定义一个停止鞅：如果边暴露过程的任意一步出现前临界复形，该鞅就停止；然后利用 Warnke 不等式证明该鞅的集中界。最终的集中界就是停止鞅的集中界，再加上该鞅曾遇到前临界复形的概率。

固定 $p = \frac{c}{n}$ 和 $m = \binom{n}{2}$ 。对于 $1 \le i \le m$， $e_i \sim Bernoulli(p)$ 是对应边是否存在的独立同分布随机变量。显然， $X(n,p) = e_1 \times \cdots \times e_m$ 是概率空间。现在，我们可以通过令 $\mathcal{F}_i$ 为对应于 $e_1, \cdots, e_i$ 的 σ-代数，在 $X(n,p)$ 上定义一个由 σ-代数组成的滤过 $\{\mathcal{F}_i\}_{i=0}^{m}$ 。

令 $X \sim X(n,p)$。现在我们构造一个边暴露鞅（例如，参见 [1] ，了解边暴露鞅的定义），如下所示：显然， $Y_m = |f_0(R_t(X))|$ 且 $Y_0 = \mathbb{E}(|f_0(R_t(X))|)$。

本节旨在证明以下集中性结果：第一阶段经过 t 轮剪枝后，复形的大小以高概率接近其期望值。

定理 11. （主定理）设 $X \sim X(n,p)$。设 $|f_0(R_t(X))|$ 为经过 t个强塌缩阶段后的顶点数，且 $Y_0 = \mathbb{E}(|f_0(R_t(X))|)$ 为其期望值。那么，对于任意 $s \ge 0$ ，有，

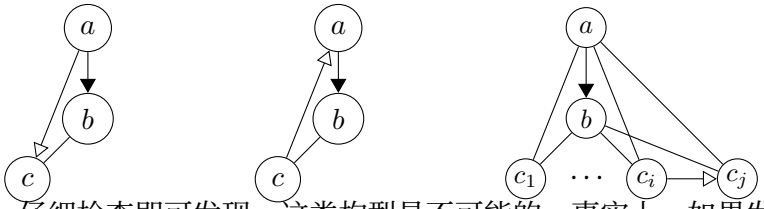
$$\begin{aligned} Pr\{||f_0(R_t(X))| - Y_0| \ge s \cdot n^{\frac{2}{3}}\} &\le \\ 2\exp\left(-\frac{s^2 \cdot n^{\frac{1}{3}}}{(c4^{t+1} + (2/3)sn^{-1/3}2^{t+1}) + O(1/n)}\right) + O(1/n) &= o_n(1). \end{aligned}$$

为证明这一点，我们先考察一些组合结果。在下面的引理中，我们将证明，删除一个顶点所产生的影响以高概率有界。

引理 12. Pr\{删除一个顶点 b 会产生 k 个新生成的被支配顶点\} $\le O(\frac{1}{n^{3k-4}})$。

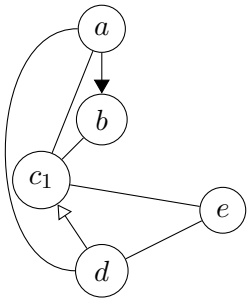
引理 *12* 的证明。 我们首先断言，删除一个顶点 b 后会产生 k 个新生成的被支配顶点，而这些顶点随后 b 至少有 k 个邻居，其中一个是 b 的支配顶点。在下图中，实线箭头表示支配，白色箭头表示

只有在删除顶点 b 后，下一阶段才会发生后续支配。箭头尖端指向被支配顶点。顶点 a， b, c 可能与其他顶点相连。下图展示了若干可能的构型。

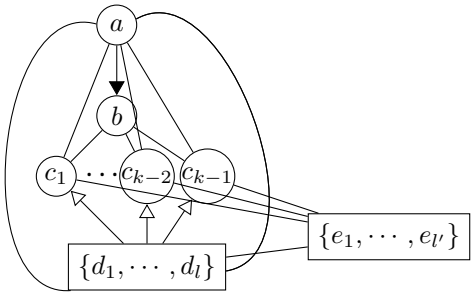


仔细检查即可发现，这类构型是不可能的。事实上，如果发生这种情况，就意味着原本将被支配的顶点已经被支配了。这是因为我们只删除了 b ，而 b 是所有原本将被支配的支配对的公共邻点。因此， b 的邻居之间不可能存在白色箭头。

因此，图1给出了 k 个新生成的被支配顶点产生所需的最小构型。在下图中，假设所有 c_i 及其对应的 d都与 $\{e_1, \cdots, e_{l'}\}$集合中某个 a 的非邻点相连。我们断言，在这种情况下，$f_1 - f_0 \geq (k-2) + (k-1) + (k-1) \geq 3k-4$。我们将对 $k \geq 2$使用归纳法证明这一断言。$k = 2$ 的情形可由下图直接看出。



现在我们证明归纳步骤。再次考虑下图。



现在假设该命题对 $k-1$ 成立。此时只需添加 kth 顶点 c_{k-1} ，即可使$f_1 - f_0$ 增加 1。同时，相应的 d 和 e 也分别使 $f_1 - f_0$ 增加 1。至此，归纳步骤完成。因此，在所有可能的最小构型中 $f_1 - f_0 \geq 3k-4$。所以，此类构型的期望数量为 $\binom{n}{f_0} \cdot (c/n)^{f_1} = O(\frac{1}{n^{3k-4}})$。因此，根据马尔可夫不等式即可得到该结果。 □

推论13。 删除一个顶点 b 会产生 3 个新生成的被支配顶点$\} \leq O(\frac{1}{n^5})$。

—

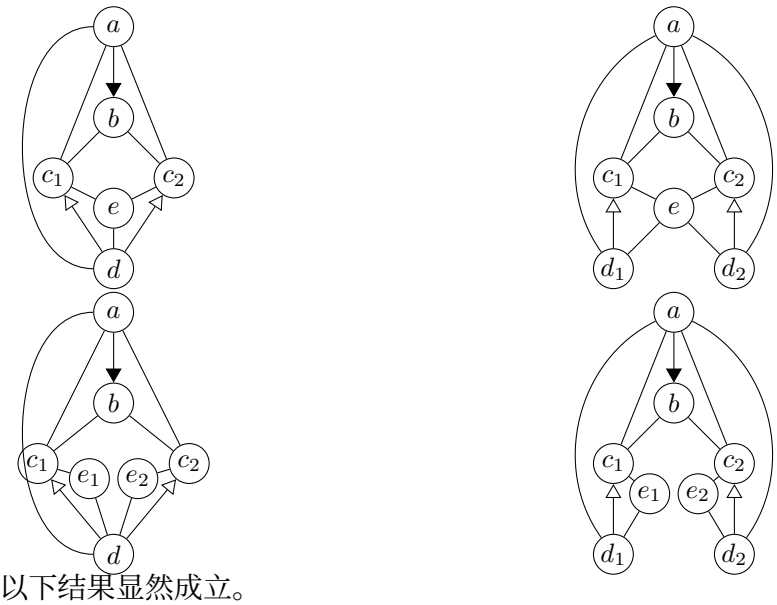
推论14。设 $X \sim X(n,\ p)$，且 $e \in f_1(X)$，则

$$Pr\{|f_0(R_t(X)) \setminus f_0(R_t(X \setminus \{e\}))| \geq 2^{t+1}\} \leq O(\frac{1}{n^5}).$$

证明。如果发生这种事件，那么过程中必定存在一个被支配顶点，删除它会产生至少3个新的支配顶点。因此，结论由前一个推论得出。

□

令临界复形 表示对应于 $k=3$ 的极小单纯复形（见下图）。令前临界复形 表示任意一个删去其中四条边的临界复形。令 N 表示 $X(n,\ p)$ 中包含临界复形的复形集合，令 N' 表示包含前临界复形的复形集合。显然， $N' \supseteq N$。



以下结果显然成立。

引理 15. 设 $X \sim X(n,p)$。则，

$$Pr\{X \in N\} \leq O(\frac{1}{n^5}),$$

and

$$Pr\{X \in N'\} \leq O(\frac{1}{n}).$$

现在在 $\{\mathcal{F}_i\}_{i=0}^m$ 上定义停止时 τ，使得若 $t=$，则 $\tau = t\ \min_{(s\leq t)}\{\mathcal{F}_s \in N'\}$。关于 $\{\mathcal{F}_i\}_{i=0}^m$，定义停止鞅 $M_i := Y_{i\wedge\tau}$。

我们先证明以下定理。

定理 16.（停止鞅不等式）设 $\{M_i\}_{i=0}^m$ 为如上定义的停止鞅。则对任意 $s \geq 0$，有：

$$Pr\{|M_m - M_0| \geq s \cdot n^{2/3}\} \leq 2\exp\left(-\frac{s^2 \cdot n^{1/3}}{(c4^{t+1} + (2/3)sn^{-1/3}2^{t+1}) + O(1/n)}\right).$$

为证明上述定理，我们将使用 Warnke [20]的以下引理。假设 $\{\mathcal{F}_K\}_{0\leq k\leq N}$是 σ-代数的递增序列，且 $\{M_K\}_{0\leq k\leq N}$是一个$\{\mathcal{F}_K\}_{0\leq k\leq N}$-适应的有界鞅。

引理 17。（有界方差鞅不等式的双侧版本）设 U_k 是一个满足 $|M_k - M_{k-1}| \leq U_k$的 F_{k-1}变量。令 $C_k =$最大值$_{i\in[k]}\, U_k$ 以及 $V_k = \sum_{i\in[k]} V\ (M_i - M_{i-1}|F_{i-1})$。令 $\phi\ (x)\ =\ (1+x)\ \log\ (1+x)\ -x$。对于每个 $s \geq 0$和 $V,\ C > 0$我们有

$$Pr\{|M_K - M_0| \geq s, V_k \leq V, C_k \leq C\ for\ some\ k \in [N]\} \leq 2e^{-s^2/(2V+2Cs/3)}$$

定理 11 基本上可由定理 16 和引理 15 推出。

证明。定理 16 的证明

令 $X \in X(n,\ p)$。我们首先尝试计算 $Pr\{X \in N|e_i,\cdots,e_1\}$，其中 $(e_1,\ \cdots,\ e_i)$ 不构成任何前临界复形。令 $M \subset N$ 为包含某个临界复形且该复形不涉及 $\{e_i,\cdots,e_1\}$ 中任何边的复形集合，并令 $M' \subset N$ 为所有临界复形都涉及 $\{e_i,\cdots,e_1\}$ 中某些边的复形集合。显然，$N = M \sqcup M'$。因此，

$$\begin{aligned} Pr\{X \in N|e_i,\cdots,e_1\} &= Pr\{X \in M \sqcup M'|e_i,\cdots,e_1\} \\ &= Pr\{X \in M|e_i,\cdots,e_1\} + Pr\{X \in M'|e_i,\cdots,e_1\} \\ &= Pr\{X \in M\} + Pr\{X \in M'|e_i,\cdots,e_1\} \end{aligned}$$

$Pr\{X \in M\}$ 是 $\{e_{i+1},\cdots,e_m\}$ 包含临界复形的概率。采用与引理 12 的证明类似的推理，可得 $Pr\{X \in M\} = O(1/n^5)$。另一方面，注意到由于 $(e_i,\ \cdots,\ e_1)$ 不包含任何前临界复形，X 至少还需要 5 条边，才能形成一个涉及 $\{e_i,\cdots,e_1\}$ 中某些边的临界复形。假设根据 $(e_1,\ \cdots,\ e_i)$，还需要 k 条边才能补全一个临界复形。显然，$5 \le k \le 13$。另外注意，在一个临界复形中，度为二的顶点至多有 2 个，其余顶点的度至少为 3。因此，即使在最坏情况下，也只需选取 3 个顶点并构造 5 条指定的边。于是 $Pr\{X \in M'|e_i,\cdots,e_1\} = O(1/n^2)$。因此，在 $(e_1,\ \cdots,\ e_i)$ 不构成任何前临界复形的条件下，$Pr\{X \in N|e_i,\cdots,e_1\} = O(1/n^2)$。特别地，

$$Pr\{|f_0(R_t(X)) \setminus f_0(R_t(X \setminus \{e_{i+1}\}))| \ge 2^{t+1}|e_i,\cdots,e_1\} \le O(\frac{1}{n^2})$$

在相同假设下。

Thus

$$\mathbb{E}(|f_0(R_t(X)) \setminus f_0(R_t(X \setminus \{e_{i+1}\}))||e_i,\cdots,e_1) \le 2^{t+1} + n \cdot O(1/n^2) \le 2^{t+1} + O(1/n)$$

当 $(e_i,\ \cdots,\ e_1)$ 不包含任何前临界复形时。

现在我们断言 $|M_{i+1} - M_i| \le 2^{t+1} + O(1/n)$。如果 $(e_i,\ \cdots,\ e_1)$ 包含一个前临界复形，则鞅停止，断言成立。现在假设 $(e_i,\ \cdots,\ e_1)$ 不包含任何前临界复形。那么

$$\begin{aligned} |[M_{i+1} - M_i](1,e_i,\cdots,e_1)| &= |M_{i+1}(1,e_i,\cdots,e_1) - \mathbb{E}_{e_{i+1}}[M_{i+1}]| \\ &= |M_{i+1}(1,e_i,\cdots,e_1) - p \cdot (M_{i+1}(1,e_i,\cdots,e_1)) - (1-p) \cdot (M_{i+1}(0,e_i,\cdots,e_1))| \\ &= |(1-p) \cdot (M_{i+1}(1,e_i,\cdots,e_1) - M_{i+1}(0,e_i,\cdots,e_1))| \\ &\le (1-p)(2^{t+1} + O(1/n)) \end{aligned}$$

类似地，

$$\begin{aligned} |[M_{i+1} - M_i](0,e_i,\cdots,e_1)| &= |M_{i+1}(0,e_i,\cdots,e_1) - \mathbb{E}_{e_{i+1}}[M_{i+1}]| \\ &= |M_{i+1}(0,e_i,\cdots,e_1) - p \cdot (M_{i+1}(1,e_i,\cdots,e_1)) - (1-p) \cdot (M_{i+1}(0,e_i,\cdots,e_1))| \\ &= |p \cdot (M_{i+1}(0,e_i,\cdots,e_1) - M_{i+1}(1,e_i,\cdots,e_1))| \\ &\le p(2^{t+1} + O(1/n)) \end{aligned}$$

因此断言得证。

接下来我们声称 $Var(M_{i+1} - M_i|\mathcal{F}_i) = Var(M_{i+1}|\mathcal{F}_i) \le (c/n)(4^{t+1} + O(1/n)) \le O(1/n)$。事实上，如果 $(e_i,\cdots,e_1)$ 包含一个前临界复形，那么鞅就会停止，且方差为零。否则

$$
\begin{aligned}
Var(M_{i+1}-M_i|e_i,\ \cdots,\ e_1) &= p\cdot([M_{i+1}-M_i](1,\ e_i,\ \cdots,\ e_1))^2+(1-p)\cdot([M_{i+1}-M_i](0,\ e_i,\ \cdots,\ e_1)\quad)^2\\
&\le p(1-p)^2(2^{t+1}+O(1/n))^2+(1-p)p^2(2^{t+1}+O(1/n))^2\\
&\le p(1-p)(4^{t+1}+O(1/n))\\
&\le (c/n)(4^{t+1}+O(1/n))
\end{aligned}
$$

因此，由引理 17,

$$
\begin{aligned}
Pr\{|M_m-M_0|\ge s\cdot n^{\frac{2}{3}}\} &\le 2\exp(-\frac{s^2\cdot n^{\frac{4}{3}}}{(n^2-n)\cdot(c/n)(4^{t+1}+O(1/n))+2/3\cdot(2^{t+1}+O(1/n))\cdot s\cdot n})\\
&\le 2\exp(-\frac{s^2\cdot n^{\frac{4}{3}}}{(n-1)c4^{t+1}+(2/3)sn^{2/3}2^{t+1}+O(1)})\\
&\le 2\exp(-\frac{s^2\cdot n^{\frac{4}{3}}}{nc4^{t+1}+(2/3)sn^{2/3}2^{t+1}+O(1)})\\
&\le 2\exp(-\frac{s^2\cdot n^{\frac{4}{3}}}{n(c4^{t+1}+(2/3)sn^{-1/3}2^{t+1})+O(1)})\\
&\le 2\exp(-\frac{s^2\cdot n^{\frac{1}{3}}}{(c4^{t+1}+(2/3)sn^{-1/3}2^{t+1})+O(1/n)})
\end{aligned}
$$

□

证明。定理 *11* 的证明 定理 11 可由定理 16 和引理 15 根据以下不等式得出。

$$
\begin{aligned}
Pr\{|Y_m-Y_0|\ge s\} &= Pr\{|Y_m-Y_0|\ge s\ and\ \neg N'\}+Pr\{|Y_m-Y_0|\ge s\ and\ N'\}\\
&\le Pr\{|M_m-M_0|\ge s\}+Pr\{N'\}
\end{aligned}
$$

Thus

$$
Pr\{|Y_m-Y_0|\ge s\cdot n^{\frac{2}{3}}\}\le 2\exp\left(-\frac{s^2\cdot n^{\frac{1}{3}}}{(c4^{t+1}+(2/3)sn^{-1/3}2^{t+1})+O(1/n)}\right)+O(1/n)=o_n(1).
$$

□

现在令 $X_0':=|f_0(R_t(X))|-|f_0(R_{t+1}(X))|$。

引理 18。对于任意 $s>0,\ \ Pr\{|X_0'-\mathbb{E}[X_0']|>sn^{\frac{2}{3}}\}<o_n(1)$。

引理 *18* 的证明。注意到

$$
\begin{aligned}
|X_0'-\mathbb{E}[X_0']|>t &\implies |(|f_0(R_{t-1}(X))|-|f_0(R_t(X))|)-(\mathbb{E}[|f_0(R_{t-1}(X))|]-\mathbb{E}[|f_0(R_t(X))|])|>t\\
&\implies ||f_0(R_t(X)|-\mathbb{E}[|f_0(R_t(X))|]|+||f_0(R_{t-1}(X)|-\mathbb{E}[|f_0(R_{t-1}(X))|]|>t\\
&\implies ||f_0(R_t(X)|-\mathbb{E}[|f_0(R_t(X))|]|>t/2\quad\text{or}\quad||f_0(R_{t-1}(X)|-\mathbb{E}[|f_0(R_{t-1}(X))|]|>t/2
\end{aligned}
$$

因此，根据并集界,

$$
\begin{aligned}
&Pr\{|X_0'-\mathbb{E}[X_0']|>sn^{\frac{2}{3}}\}\\
\le\ &Pr\{||f_0(R_t(X)|-\mathbb{E}[|f_0(R_t(X))|]|>(s/2)n^{\frac{2}{3}}\}+Pr\{||f_0(R_{t-1}(X)|-\mathbb{E}[|f_0(R_{t-1}(X))|]|>(s/2)n^{\frac{2}{3}}\}\le o_n(1)
\end{aligned}
$$

□

令 X_0 表示第一阶段结束时被支配顶点数量的随机变量。显然 $0\le X_0\le|f_0(R_t(X))|-|f_0(R_{t+1}(X))|=X_0'$。由于 $X_0'\le\mathbb{E}[X_0']+o(n)$ 几乎必然成立，我们得到 $0\le X_0\le\mathbb{E}[|f_0(R_t(X))|]-\mathbb{E}[|f_0(R_{t+1}(X))|]+o(n)$ 几乎必然成立。

7 第二阶段

第二阶段将是第一阶段的一个较慢版本。在此阶段，均匀随机地选取一个被支配顶点并将其移除。该过程持续进行，直到不再存在被支配顶点。与 [3], 的证明类似，我们将证明：删除一个被支配顶点时，新产生的被支配顶点的期望数量严格小于 1，因此在 $o(n)$ 步之内，强塌缩过程将停止。因此，核的大小将等于第一阶段结束后剩余顶点的数量，再乘以一个 $o(n)$ 因子（忽略该因子）。

设经过 t 个剪枝阶段后，第一阶段结束，第二阶段开始。另设 Y_i 为随机变量，表示在第二阶段第 i 步中，仅由删除被支配顶点而新生成的被支配顶点的数量。注意，$Y_i \in \{0, \cdots, n\}$。首先，我们尝试计算 $Pr\{Y_i = 1\}$。

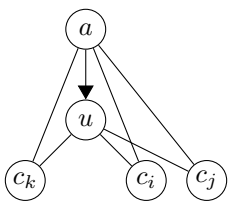
引理 19. 对于任意 $i \leq \frac{1}{12}n(1-\gamma)(1-c\gamma)$，有 $\mathbb{E}[Y_i] \leq 1 - \frac{3}{4}(1-c\gamma) + o_n(1) < 1 + o_n(1)$。

引理 19 的证明。 如果某个顶点的度在第二阶段的第 i 次坍缩中发生变化，我们就称该顶点受到这次坍缩的影响。定义 Q_i 为事件 $\{Y_i = 1\}$ 的子集，其中，新生成的顶点首次受到第 i 次坍缩的影响。显然，事件$\{\{Y_i = 1\} \cap Q_i\}$ 表示第 i 次坍缩（删除顶点 u）仅新生成了一个顶点，比如 v，且 v 首次受到影响。因此，该特定顶点保留了自第一阶段以来的局部结构。这里的思路是，第一阶段结束后，边 $\{u, v\}$ 可以连接到任何可能的位置。我们只计算它连接到 $R_t(X) \setminus \{\{u, v\}, \{u\}\}$ 中某个合适顶点的概率。为了计算 $Pr\{\{Y_i = 1\} \cap Q_i\}$，我们将其进一步划分为两个事件。为此，定义 P 为事件：经过 $i-1$ 步后，$deg(v) = 2$。因此，概率 $Pr\{\{Y_i = 1\} \cap Q_i \cap P\}$ 等于 $R_t(X) \setminus \{\{u, v\}, \{u\}\}$ 中度为 1 的顶点数量与第一阶段经过 $t-1$ 阶段后非孤立顶点数量之比，计算方式与 [3] 的式 8 相同。利用与第 6 节类似的论证可以证明，这两个量都集中在各自的均值附近。这两个量分别为 $c(1-\gamma_t)\gamma_{t+1}n + o(n)$ 和 $(1-\gamma_t)n + o(n)$，且该结论以渐近概率 1 成立。

要使事件 $\{Y_i = 1\} \cap Q_i \cap P$ 发生，v 必须是某个 2-单形的一部分。现在我们来界定第一阶段结束后剩余的 2-单形数量。注意，在塌缩阶段，2-单形的数量只会减少。令随机变量$T := |f_2(X)|$ 和 $T' := |f_2(R_t(X))|$ 表示 $X \sim X(n, c/n)$。显然，$T' \leq T$。由马尔可夫不等式可得 $Pr\{T > \log(n)\} \leq O(1/\log(n))$。因此，$Pr\{T' > \log(n)\} \leq O(1/\log(n))$。于是 $Pr\{\{Y_i = 1\} \cap Q_i \cap P\} \leq O(\log(n))/(1-\gamma_t)n$ 几乎必然成立。因此，几乎必然有 $Pr\{\{Y_i = 1\} \cap Q_i\} \leq c(1-\gamma_t)\gamma_{t+1}n\,(1-\gamma_t)n \pm O(\log(n))/(1-\gamma_t)n \leq c(1-\gamma_t)\gamma_{t+1}\,(1-\gamma_t) + o_n(1)$。

现在我们需要计算 $Pr\{\{Y_i = 1\} \cap Q_i\}$。为此，我们再次将该事件划分为两个互不相交的事件。令 A_i 表示第二阶段第 i 步中受影响顶点的数量。现在定义事件 $B_i := \bigcap_{j=1}^{i}\{A_j \leq 3\}$。因此，$Pr\{\{Y_i = 1\} \cap Q_i \cap B_{i-1}\} \leq 3\,(i-1)\,n\,(1-\gamma_t)$。

为了计算 $Pr\{\{Y_i = 1\} \cap Q_i \cap \overline{B_{i-1}}\}$，首先注意到 $Pr\{\{Y_i = 1\} \cap Q_i \cap \overline{B_{i-1}}\} \leq Pr\{B_{i-1}\} \leq \Sigma_{j=1}^{i-1}\{A_j \geq 4\}$。但要使 $\{A_i \geq 4\}$ 发生，相应的被支配顶点 u必须属于以下排列。



令 S 和 S' 分别表示 X 和 $R_t(X)$ 中此类构型的数量，对于 $X \sim X(n, c/n)$。显然，$S' \leq S$。由马尔可夫不等式可得 $Pr\{S \geq 1\} \leq O(1/n^2)$。因此，$Pr\{S' \geq 1\} \leq O(1/n^2)$。所以，渐近几乎必然不会发生 $\{B_i\}$。

类似的论证结合引理 12 可得 $Pr\{Y_i \in \{2, \cdots, n\}\} \leq O\left(1/n^2\right)$。收集所有项后，我们得到如下不等式。

$$
\begin{aligned}
\mathbb{E}[Y_i] &= \Sigma_{j=0}^{n} j \cdot Pr\{Y_i = j\} \\
&\leq Pr\{Y_i = 1\} + n \cdot O(1/n^2) \\
&\leq Pr\{\{Y_i = 1\} \cap Q_i\} + Pr\{\{Y_i = 1\} \cap \overline{Q_i}\} + n \cdot O(1/n^2) \\
&\leq Pr\{\{Y_i = 1\} \cap Q_i\} + Pr\{\{Y_i = 1\} \cap \overline{Q_i} \cap B_{i-1}\} + Pr\{\{Y_i = 1\} \cap \overline{Q_i} \cap \overline{B_{i-1}}\} + n \cdot O(1/n^2) \\
&\leq \frac{c(1-\gamma_t)\gamma_{t+1}}{1-\gamma_t} + \frac{3(i-1)}{n(1-\gamma_t)} + O(1/n) + o_n(1) \\
&\leq c\gamma_{t+1} + \frac{3(i-1)}{n(1-\gamma)} + o_n(1) \\
&\leq c\gamma + \epsilon_i + o_n(1) \\
&\leq c\gamma + \frac{1}{4}(1 - c\gamma) + o_n(1) \\
&\leq 1 - \frac{3}{4}(1 - c\gamma) + o_n(1) < 1 + o_n(1), \text{ by lemma 5.}
\end{aligned}
$$

□

注意，对于任意固定的 i, $\epsilon_i = o_n(1)$，以及 $c > 1$，我们有 $c\gamma < 1$。

令 X_i 为第二阶段第 i 步结束时被支配顶点的数量。则有

$$X_i = X_{i-1} - 1 + Y_i = X_0 - n + \Sigma_{i=1}^{n} Y_i$$

直到第二阶段结束。如果第二阶段在第 i' 步停止，则 $X_j = 0 \quad \forall j > i'$。因此我们有

$$
\begin{aligned}
\mathbb{E}[X_0] &\leq \mathbb{E}[|f_0(R_t(X))|] - \mathbb{E}[|f_0(R_{t+1}(X))|] \\
&= (1 - \gamma_{t+1} - c\gamma_t + c\gamma_t^2)n - (1 - \gamma_{t+2} - c\gamma_{t+1} + c\gamma_{t+1}^2)n
\end{aligned}
$$

and,

$$\mathbb{E}[X_i] = \mathbb{E}[X_{i-1}] - 1 + Y_i = \mathbb{E}[X_0] - n + \Sigma_{i=1}^{n} Y_i$$

只要第二阶段继续进行。

现在定义 $\delta \equiv \delta(t) := (1 - \gamma_{t+1} - c\gamma_t + c\gamma_t^2) - (1 - \gamma_{t+2} - c\gamma_{t+1} + c\gamma_{t+1}^2)$，使得 $\mathbb{E}[X_0] \leq n\delta$。

现在我们给出本节的主要引理。下面的第一个引理表明，以高概率，对于任意足够小的 $\varepsilon > 0$，我们都可以选择一个足够大的 t，使得第二阶段删除的顶点数量小于 εn。证明方法是将该阶段经过 i 步后剩余的被支配顶点数量建模为偏置随机游走。

引理 20. $\forall 0 < \epsilon < \min\{\frac{1}{12}(1-\gamma)(1-c\gamma), \frac{5}{192}(1-\gamma)(1-c\gamma)^2\}$ $\exists T$使得 $\forall t > T$几乎必然至多有 ϵn 个顶点会在算法到达核之前从 $R_t(X)$ 中被删除。

引理 *20* 的证明. 选择 t 使得 $\delta \equiv \delta(t) < \frac{1}{8}\epsilon(1 - c\gamma)$。这是可行的，因为随着 t增大， $\gamma_t - \gamma_{t+1}$ 趋于 $\gamma_{t+1} - \gamma_{t+2}$ 零。

现在假设第二阶段持续 $\rho = \epsilon n$ 步。我们将证明 $\mathbb{E}[X_\rho] = 0$几乎必然成立，即不再有可删除的被支配顶点。为此，我们定义如下新随机变量序列 $\{Z_i\}$：

$$Z_0 := X_0$$

and,

$$Z_i := Z_{i-1} - 1 + Y_i = Z_0 - n + \Sigma_{i=1}^{n} Y_i$$

注意，$X_i \le Z_i$ 和 $Z_i \le 0 \implies X_i = 0$。

由于 $\rho = \epsilon n \le \frac{1}{4}n(1-\gamma)(1-c\gamma)$，在经过 ρ 步后剩余的被支配顶点数量为

$$\begin{aligned}
\mathbb{E}[X_\rho] \le \mathbb{E}[Z_\rho] &= \mathbb{E}[Z_0] - \Sigma_i^\rho 1 + \Sigma_i^\rho \mathbb{E}[Y_i] \\
&\le n\delta - \frac{3}{4}\rho(1-c\gamma) && \text{根据引理 19} \\
&\le \frac{1}{8}n\epsilon(1-c\gamma) - \frac{3}{4}n\epsilon(1-c\gamma) \\
&\le -\frac{5}{96}(1-\gamma)(1-c\gamma)^2 n \le 0
\end{aligned}$$

因此 $\mathbb{E}[X_\rho] = 0$。

现在我们将证明集中性。定义 $l := \frac{5}{96}(1-\gamma)(1-c\gamma)^2$。接下来我们证明

$$Pr\{Z_\rho > 0\} \le Pr\{Z_\rho - \mathbb{E}[Z_\rho] > l \cdot n\} < o_n(1)$$

将 Z_ρ 写成 $Z_\rho \equiv Z_0 + Z'_\rho(Y_1, \cdots, Y_\rho)$。首先注意到，根据引理 18，$Pr\{Z_0 - \mathbb{E}[Z_0] > (l/2)n\} \le Pr\{|Z_0 - \mathbb{E}[Z_0]| > (l/2)n\} \le o_n(1)$ 对任意 $s > 0$ 都成立。

由主要几何引理可得

$$\mathbb{E}[e^{t(Y_i - \mathbb{E}[Y_i])}] \le \mathbb{E}[e^{tY_i}] \le 1 + e^t + O(\frac{e^{2t}}{n^2} \cdot \frac{1-(e^t/n^3)^{n-1}}{1-e^t/n^3})$$

固定 $t > 0$。然后，

$$\begin{aligned}
Pr\{Z'_\rho - \mathbb{E}[Z'_\rho] > (l/2)\cdot n\} &= Pr\{\Sigma_i^\rho Y_i - \Sigma_i^\rho \mathbb{E}[Y_i] > (l/2)\cdot n\} \\
&= Pr\{e^{t(\Sigma_i^\rho Y_i - \Sigma_i^\rho \mathbb{E}[Y_i])} > e^{tn(l/2)}\} \\
&\le e^{t(\Sigma_i^\rho Y_i - \Sigma_i^\rho \mathbb{E}[Y_i])} / e^{tn(l/2)} \\
&\le \frac{(1 + e^t + O(\frac{e^{2t}}{n^2} \cdot \frac{1-(e^t/n^3)^{n-1}}{1-e^t/n^3}))^\rho}{e^{tn(l/2)}}
\end{aligned}$$

令 $t = \log(n)$，我们得到

$$\begin{aligned}
Pr\{Z'_\rho - \mathbb{E}[Z'_\rho] > (l/2)\cdot n\} &\le \frac{(1 + n + O(\frac{1-(1/n^2)^{n-1}}{1-1/n^2}))^\rho}{n^{n(l/2)}} \\
&\le \frac{(1 + n + O(\frac{1}{1-1/n^2}))^{\epsilon n}}{n^{n(l/2)}} \\
&\le \frac{((O(1)+n)^\epsilon)^n}{(n^{(l/2)})^n} \\
&\le (\frac{(O(1)+n)^\epsilon}{n^{(l/2)}})^n \le o_n(1)
\end{aligned}$$

由于 $\epsilon < l/2$，该量随着 n 的增大而趋于零，因此最后一个不等式成立。所以，

$$Pr\{Z_\rho - \mathbb{E}[Z_\rho] > l\cdot n\} \le Pr\{Z_0 - \mathbb{E}[Z_0] > (l/2)\cdot n\} + Pr\{Z'_\rho - \mathbb{E}[Z'_\rho] > (l/2)\cdot n\} \le o_n(1)$$

□

下一个引理由 γ_t 的性质以及第 6 节中给出的集中界推得。

引理 21. $\forall 0 < \delta \quad \exists T$ 使得 $\forall t > T$ 几乎必然成立。

$$(1-\gamma)(1-c\gamma)n + o(n) \le |f_0(R_t(X))| \le (1-\gamma)(1-c\gamma)n + \delta n + o(n).$$

引理 *21* 的证明。 由定理 6 和定理 11 可知，几乎必然有

$$|f_0(R_t(X))| = (1-\gamma_{t+1} - c\gamma_t + c\gamma_t^2)n + o(n)$$

定义 $h(x) := 1 - e^{-c(1-x)} - cx(1-x)$，其定义域为区间 (0,1)。容易验证，在此区间上 $h'(x) < 0$，即 $h(x)$ 严格递减。因此，我们有 $(1-\gamma_{t+2} - c\gamma_{t+1} + c\gamma_{t+1}^2) \le (1-\gamma_{t+1} - c\gamma_t + c\gamma_t^2)$。由此可得左侧不等式。

注意， $\{\gamma_t\}_t$ 是一个单调递增且收敛于 γ的序列。因此，$\{(1-\gamma_{t+1} - c\gamma_t + c\gamma_t^2)\}_t$是一个单调递减且收敛于 $(1-\gamma-c\gamma+c\gamma^2)$的序列。由此，通过取充分大的 T，对于任意 $\delta > 0$，我们可以将 $\{(1-\gamma_{t+1} - c\gamma_t + c\gamma_t^2)\}_{t>T}$限制在以 $(1-\gamma-c\gamma+c\gamma^2)$ 为中心的一个 δ-球内。由此可得右侧不等式。 □

利用上述两个引理，我们得到第一个主要结果定理 1 的证明，该结果讨论了 Erdős–Rényi 复形强塌缩后核的大小。

定理 *1* 的证明。由引理 20 和引理 21 可得。 □

8 端点相变

令 $P(X)$ 表示所有可能的被支配—支配顶点对的数量。可以证明，对于 $X \sim X(n,p)$， $\mathbb{E}[P(X)] = (n(n-1)/2)p(1-p(1-p))^{n-2}$。对于 $p = \frac{\lambda}{\log(n)\, n}$，其中 $\lambda > 1$,$\mathbb{E}[P(X)] = O(\frac{n}{\log(n)\, n^{\lambda}}) = o_n(1)$。因此，根据马尔可夫不等式，几乎必然不存在可启动坍缩过程的被支配顶点。

现在我们将关注 $X \sim X(n,p)$ 在 $p = 1 - \frac{\lambda}{\log(n)\, n}$ 时的行为。对于 $\lambda > 2$,$\mathbb{E}[P(X)] = O(\frac{n^2}{n^{\lambda}}) = o_n(1)$。因此，几乎必然不存在可启动坍缩过程的被支配顶点。但正如下面的引理所述，当 $\lambda < 1$ 时，情况则截然相反。

引理 22。 对于 $X \sim X$（X、$1 - \frac{\lambda}{\log n}$）和 $\lambda < 1$， X几乎必然是可坍缩的。

引理 *22* 的证明。 我们将证明，在此范围内，几乎必然存在一个与其他所有顶点都相邻的顶点。定义随机变量 $V \in \{0, \cdots, n\}$，其统计与所有其他顶点都相邻的顶点数量。显然，$\mathbb{E}[V] = np^{n-1} = n(1-\frac{\lambda}{\log(n)\ n})^{n-1} = \Theta(\frac{n}{n^{\lambda}})$。现在我们来计算 $Var(V)$。令 I_i 表示指示随机变量，该变量在 v_i 与所有其他顶点都相邻时取值为 1。于是，

$$\begin{aligned} Var(V) &= nVar(I_1) + n(n-1)cov(I_1, I_2) \\ &= np^{n-1}(1-p^{n-1}) + n(n-1)(p^{2n-3} - p^{2n-2}) \end{aligned}$$

因此，

$$\begin{aligned} Pr\{V=0\} &\le \frac{Var(V)}{(\mathbb{E}[V])^2} \\ &\le \frac{np^{n-1}(1-p^{n-1}) + n(n-1)(p^{2n-3} - p^{2n-2})}{n^2 p^{2n-2}} \\ &\le 1/(np^{n-1}) + (1/p - 1) \\ &\le \Theta(\frac{n^{\lambda}}{n}) + \frac{\lambda \log(n)}{n - \lambda \log(n)} = o_n(1) \end{aligned}$$

因此 $V \geq 1$ 几乎必然成立。

□

参考文献

[1] N. Alon 和 J. Spencer。《概率方法》。Wiley，纽约，第3版，2008年。[2] Lior Aronshtam 和 Nathan Linial。随机单纯复形的顶维同调何时消失？《*Random Struct. Algorithms*》，46(1):26–35，2015年。doi:10.1002/rsa.20495。[3] Lior Aronshtam 和 Nathan Linial。随机复形中 d-可塌缩性的阈值。《*Random Struct. Algorithms*》，48(2):260–269，2016年。doi:10.1002/rsa.20585。[4] Lior Aronshtam、Nathan Linial、Tomasz Luczak 和 Roy Meshulam。随机单纯复形的可塌缩性与顶维同调消失。《*Discret. Comput. Geom.*》，49(2):317–334，2013年。doi:10.1007/s00454-012-9483-8。[5] Dominique Attali、Andr´e Lieutier 和 David Salinas。Vietoris–Rips复形也能对采样形状进行拓扑正确重建。《计算几何》，46(4):448–465，2013年。[6] J. A. Barmak 和 E. G. Minian。强同伦类型、神经与塌缩。《离散与计算几何》，47:301–328，2012年。[7] J-D. Boissonnat 和 S. Pritam。通过强塌缩计算旗复形的持久同调。《国际计算几何研讨会（*SoCG*）》，2019年。[8] J-D. Boissonnat 和 S. Pritam。旗复形的边塌缩与持久性。《国际计算几何研讨会（*SoCG*）》，2020年。[9] J-D. Boissonnat、S.Pritam 和 D. Pareek。《持久性的强塌缩》。载于《第*26*届欧洲算法研讨会（*ESA 2018*）》，第112卷，2018年。[10] B´ela Bollob´as。《随机图》。剑桥高等数学研究丛书第73卷。剑桥大学出版社，第2版，2001年。[11] David A. Freedman。鞅的尾概率。《概率年刊》，3(1):100–118，1975年。[12] Marc Glisse 和 Siddharth Pritam。交换、移位和修剪以实现滤过的边塌缩。载于《第*38*届国际计算几何研讨会（*SoCG 2022*）》，第224卷，页44:1–44:15，2022年。[13] A. Hatcher。《代数拓扑》。剑桥大学出版社，2001年。[14] Matthew Kahle。随机单纯复形，2016年。arXiv:1607.07069。[15] DMITRY N. KOZLOV。随机 d-复形顶维同调群消失的阈值函数。《美国数学会会刊》，138(12):4517–4527，2010年。URL: http://www.jstor.org/stable/41059187。[16] Nathan Linial 和 Roy Meshulam。随机2-复形的同调连通性。《*Comb.*》，26(4):475–487，2006年。URL: https://doi.org/10.1007/s00493-006-0027-9，doi: 10.1007/s00493-006-0027-9。[17] Nathan Linial 和 Yuval Peled。随机单纯复形：相变附近。《离散数学之旅》，页543–570，2017年。

[18]Greg Malen。随机团复形的可缩性。离散数学，346(3):113267，2023。URL：https://www.sciencedirect.com/science/article/pii/ S0012365X22004733, doi:https://doi.org/10.1016/j.disc.2022.113267。[19] Roy Meshulam 和 N. Wallach。随机 k维复形的同调连通性。*Random Struct. Algorithms*，34(3):408–417，2009。URL：https://doi.org/10. 1002/rsa.20238，doi:10.1002/rsa.20238。[20] LUTZ WARNKE。论典型有界差分方法。组合学、概率与计算，25(2):269–299，2016。doi:10.1017/S0963548315000103。[21] J. H. C Whitehead。单纯空间、核与 m 群。*Proc. London Math. Soc*，45:243–327，1939。[22] A. C. Wilkerson、H. Chintakunta 和 H. Krim。大数据中的持久特征计算：一种分布式降维方法。载于国际声学、语音与信号处理会议 *(ICASSP)*，第 11–15 页，2014。[23] Siddharth Pritam ´Angel Javier Alonso, Michael Kerber。双滤过图中的滤过支配。载于*SIAM* 算法工程与实验研讨会 *(ALENEX23)*，2023。