稀疏 Erd ˝os-R ´enyi 随机图 $G(n,m)$ 和 $G(n,p)$ 中三角形计数的中等偏差

JOS ´E D. ALVARADO、LEONARDO GONC¸ALVES DE OLIVEIRA 和 SIMON GRIFFITH S

arXiv:2305.04326v2 [math.CO] 9 Jan 2025

Abstract. 我们研究了在密度大于 $n^{-1/2}(\log n)^{1/2}$ 的 Erd˝os-R´enyi 随机图 $G(n,m)$ 和 $G(n,p)$ 中确定三角形计数偏差概率的问题。具体而言，对于 $G(n,m)$ 和 $G(n,p)$ 模型，我们在几乎整个可能偏差的范围内，将对数概率 $\log \mathbb{P}\left(N_{\triangle}(G) > (1+\delta)p^3n^3\right)$ 确定到一个常数因子以内。对于 $G(n,p)$ 模型，我们还在非局部化区域内，将结果精确到一个 $(1+o(1))$ 因子以内，从而证明了更强的结论。我们还得到了关于下尾以及樱桃（长度为 2 的路径）计数的一些结果。

1. Introduction

随机图 $G(n,p)$ 中三角形计数偏差的问题在近几十年来得到了广泛研究。我们将在第 1.2 节概述相关文献。大多数结果关注偏差的某个特定范围，例如大偏差（量级与均值相当）；而我们将证明的结果基本适用于可能偏差的整个范围。

我们也可以在 $G(n,m)$ 随机图模型中研究这些问题。在该模型中，随机图 $G \sim G(n,m)$ 从所有具有 n 个顶点和 m 条边的图中均匀选取。从概念上说，采用 $G(n,m)$ 模型的优势在于，我们排除了一种可能导致偏差的原因——边数的波动。

采用 $G(n,m)$ 模型还有实际上的理由——我们可以通过逐条揭示边、经过固定的 m 步来生成$G(n,m)$，这尤其适合运用鞅技巧。此外，在 $G(n,m)$ 中证明的结果可以很容易地转化为 $G(n,p)$ 中的结果。

那么，我们先讨论 $G(n,m)$ 中的三角形计数偏差。在陈述了这一情形下的结果后，我们将在第 1.2 节回到 $G(n,p)$ 情形。

Jos´e 曾获得巴西资助机构 CNPq（项目编号：153903/2018-0）、FAPERJ（项目编号：E-26/202.011/2020）和 FAPESP（项目编号：2020/10796-0）提供的博士后资助，在里约热内卢天主教大学 Simon 以及 USP 的 Guilherme Mota 指导下开展研究；Leonardo 获得了高等教育人员提升协调会提供的博士资助，Simon 则获得了 CNPq（项目编号：307521/2019-2）和 FAPERJ（项目编号：E-26/202.713/2018 及 E-26/201.194/2022）提供的研究资助。

1.1. 我们关于 $G(n,m)$ 中三角形计数的结果。我们使用记号 $N := \binom{n}{2}$，并设定 $t := m/N$。注意，t 是 $G_m \sim G(n,m)$ 中实际存在的边所占的比例。我们使用下降阶乘记号 $(n)_k = n(n-1)\dots(n-k+1)$。

我们注意到，我们的方法与 Goldschmidt、Scott 和第三作者 [20]的方法类似。然而，我们强调，我们引入了新的思想和技术，这些思想和技术对于将 [20] 中证明的结果推广到密度范围更广的情形是必不可少的。

我们用 $N_\triangle(G_m)$ 表示 G_m 中三角形的同构副本数量。例如 $N_\triangle(K_4) = 24$。我们定义

$$L_\triangle(m) := \mathbb{E}\left[N_\triangle(G_m)\right] = \frac{(m)_3(n)_3}{(N)_3},$$

因此

$$D_\triangle(G_m) := N_\triangle(G_m) - L_\triangle(m)$$

是该三角形计数相对于其均值的偏差。

我们在 $G(n,m)$ 模型中的主要结果，是在几乎整个偏差范围内，将与偏差相关的速率确定到一个常数乘性因子以内，对于所有密度 $t \gg n^{-1/2}(\log n)^{1/2}$。也就是说，给定一个满足 $t^{3/2}n^{3/2}\sqrt{\log n} \ll a_n \ll t^{3/2}n^3$ 的序列$a = a_n$，我们找到一个序列 $b = b_n$，使得$\mathbb{P}(D_\triangle(G_m) > a) = \exp(-\Theta(b)$。

事实上，我们发现，用概率为 $\exp(-b)$ 的偏差 a 来表述这些结果更直观。显然，这与前述表述等价。令 $\mathrm{DEV}_\triangle(b,t)$ 表示概率仅为 e^{-b} 的三角形计数偏差。也就是说，$\mathrm{DEV}_\triangle(b,t)$ 是满足以下条件的 a 的最小值：

$$\mathbb{P}\left(N_\triangle(G_m) > \mathbb{E}\left[N_\triangle(G_m)\right] + a\right) \leqslant e^{-b}.$$

俗话说，一图胜千言。因此，我们先展示一幅图，再更正式地陈述结果。特别是，这幅图将突出显示四个不同的区间，它们分别对应三角形计数偏差的不同潜在“成因”。

我们可以将偏差参数化为 a as $\delta t^3 n^3$，这在实际中约为均值的 δ 倍。该图考察了形式为 $t = n^\gamma$ 的密度 t，以及形式为$\delta = n^\theta$的 δ。对于每个 $\gamma \in (-1/2, 0)$，我们得到了介于标准差的数量级（数量级为 $t^{3/2}n^{3/2}$）与最大可能偏差的数量级（数量级为 $t^{3/2}n^3$）之间的偏差结果。这解释了 $\theta \in (-3/2 - 3\gamma/2, -3\gamma/2)$ 的取值区间。

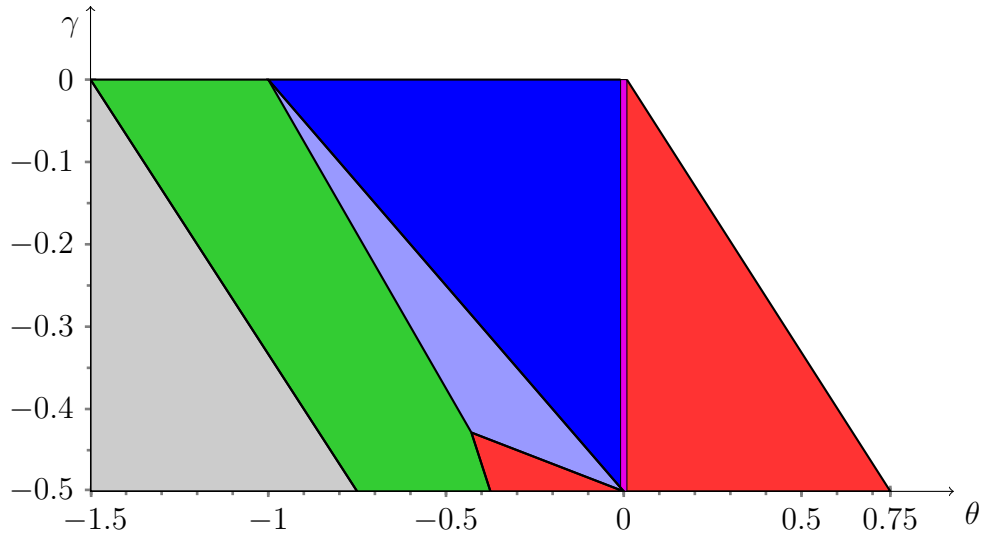


FIGURE 1. 在这幅图中，灰色表示非常小的偏差，紫色线（$\theta = 0$）对应传统的大偏差结果。其他颜色分别代表相应偏差的一种不同的“最可能成因”。绿色区域：纯粹的好运，没有结构性成因。浅蓝色区域：星形图。深蓝色区域：枢纽。红色区域：团。

现在我们来讨论偏差的四种可能成因。本文中，令 $\ell := \log(1/t)$ 表示 $1/t$ 的自然对数。

纯粹的好运（NORMAL）：证明基于一个与过程中（三角形随边的添加而生成）相关的鞅。鞅各增量的条件方差之和很可能处于 t^3n^3 的数量级。由此（利用 Freedman 不等式）可知，数量级为

$$\mathrm{NORMAL}(b,t) := b^{1/2}t^{3/2}n^{3/2}$$

的偏差发生的概率为 e^{-b}。这种行为是中心极限定理以及标准差数量级偏差行为的推广。因此，我们称之为“正态”区间。

星形图：注意到度特别大的顶点 d 很可能包含在 $\Theta(d^2t)$ 个三角形中。因此，度为 $d \gg tn$ 的星形图会“导致” $\Theta(d^2t)$ 个三角形，其出现概率约为 $t^d = e^{-d\ell}$。显然，这只适用于度 $d < n$。我们看到，这可能导致量级为

$$\mathrm{STAR}(b,t) := \frac{b^2t}{\ell^2}\,1_{b\leqslant n\ell}$$

概率为 e^{-b}。

枢纽：枢纽是星形图达到满度后的一种自然延伸。注意，度数为 n 量级的 k 个顶点将参与 $\Theta(ktn^2)$ 个三角形，而这种情况发生的概率约为 $t^{kn} = e^{-kn\ell}$。我们看到，这可能导致量级为...的偏差

$$\mathrm{HUB}(b,t) \;:=\; \frac{btn}{\ell}\,1_{b\geqslant n\ell}$$

以概率 e^{-b}。

一个团：由 k 个顶点构成的团会产生 $\Theta(k^3)$ 个三角形，其出现的概率约为 $t^{k^2} = e^{-k^2\ell}$。我们看到，这可能导致阶为

$$\mathrm{CLIQUE}(b,t) \;:=\; \frac{b^{3/2}}{\ell^{3/2}}$$

以概率 e^{-b}。

令 $M(b,t)$ 定义为这四者中的最大值，

$$M(b,t) \;=\; \max\{\mathrm{NORMAL}(b,t), \mathrm{STAR}(b,t), \mathrm{HUB}(b,t), \mathrm{CLIQUE}(b,t)\}\,. \tag{1.1}$$

我们关于 $G(n,m)$ 模型的主要定理表明，发生概率为 e^{-b} 的偏差 $\mathrm{DEV}_{\triangle}(b,t)$，在 t 和 b 的较大范围内，等于 $M(b,t)$，相差一个乘法常数。注意，我们将 t 限制为至多 $1/2$，尽管在密度与 1 保持有界距离的所有情形下，结果显然都成立，见 [20]。

定理 1.1. 存在绝对常数 c, C ，使得以下结论成立。对于所有 $Cn^{-1/2}(\log\ n)^{1/2} \leqslant t \leqslant 1/2$ 和 $3\log\ n \leqslant b \leqslant tn^2\ell$，我们有

$$cM(b,t) \;\leqslant\; DEV_{\triangle}(b,t) \;\leqslant\; CM(b,t)\,.$$

注 1.2. 在稀疏随机图中，偏差的表现可能有所不同，其中 $t \ll n^{-1/2}$，因为在这种情况下，大多数边都不属于任何三角形。这就是所谓的泊松区间。关于该区间内的大偏差结果，参见 [21] ；关于常数平均度这一更稀疏区间内的结果，参见 [19] 。我们的结果似乎很可能可以推广到 $t = Cn^{-1/2}$。难点在于，我们的结果基于共度的表现。当期望共度约为 t^2n 时，如果 t^2n 只是一个常数，就更难证明集中性结果。

这里还有一幅图，或许有助于我们直观地理解这个结果。这里我们设定$t = n^{\gamma}$ 和 $b = n^{\eta}$。

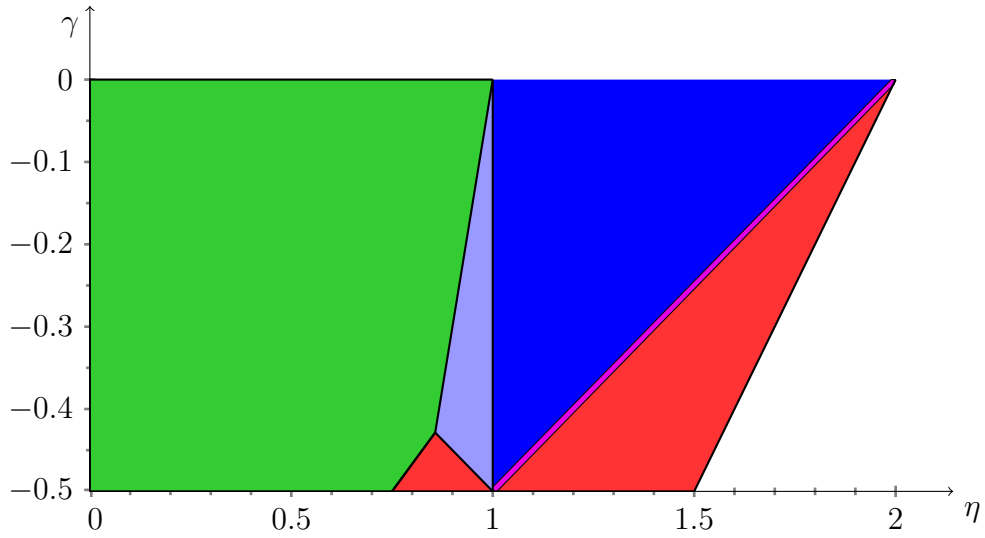


FIGURE 2. 与前一幅图一样，紫色曲线对应传统的大偏差结果。其他每种颜色分别代表导致相应偏差的一种不同的“最可能原因”。在绿色区域：没有结构性原因的好运气。在浅蓝色区域：星形图。在深蓝色区域：枢纽。在红色区域：团。

有些读者可能更喜欢将概率表示为偏差的函数的结果版本，因此我们将其作为定理 1.1 的直接推论陈述。我们定义

$$r(t,a) := \min\left\{\frac{a^2}{t^3n^3}, \frac{a^{1/2}\ell}{t^{1/2}} + \frac{a\ell}{tn}, a^{2/3}\ell\right\}.$$

推论 1.3. 存在绝对常数 c, C，使得以下结论成立。对于所有满足 $Cn^{-1/2}(\log n)^{1/2} \leqslant t \leqslant 1/2$ 和 $Ct^{3/2}n^{3/2}\sqrt{\log n} \leqslant a \leqslant ct^{3/2}n^3$ 的情形，我们有

$$\exp(-Cr(t,a)) \leqslant \mathbb{P}\left(D_\triangle(G_m) > a\right) \leqslant \exp(-cr(t,a)).$$

我们还提到，我们的方法为下尾部提供了一些结果。在讨论下尾部时，更难谈论偏差的“原因”。尽管 Neeman、Radin 和 Sadun [37] 已在这一方向取得了一些进展。在稠密情形下，当 t 为常数时，他们表明下尾部与邻接矩阵的最小特征值关系更为密切。我们的方法实际上无法处理这类结构，因此我们的下尾部结果仅限于正态区间。

定理 1.4. 存在绝对常数 c, C，使得以下结论成立。对于所有满足 $Cn^{-1/2}(\log n)^{1/2} \leqslant t \leqslant 1/2$ 和 $Ct^{3/2}n^{3/2}\sqrt{\log n} \leqslant a \leqslant ct^{3/2}n^2$ 的情形，我们有

$$\exp\left(\frac{-Ca^2}{t^3n^3}\right) \leqslant \mathbb{P}\left(D_\triangle(G_m) < -a\right) \leqslant \exp\left(\frac{-ca^2}{t^3n^3}\right).$$

1.2. $G(n,p)$ 中三角形计数的结果。近几十年来，理解 Erd˝os-R´enyi 随机图 $G(n,p)$ 中三角形数量的上尾行为，一直是概率组合学的研究重点。我们使用以下记号表示与超过均值 $a = \delta p^3(n)_3$ 的偏差相对应的渐近速率：

$$r(\delta,p,n) \;:=\; -\log \mathbb{P}\left(N_{\triangle}(G_p) \;>\; (1+\delta)p^3(n)_3\right)\,,$$

其中 $G_p \sim G(n,p)$。Vu [41] 以及 Janson 和 Ruci´nski [29] 在 21 世纪初研究了这一渐近速率，并提出了多种用于界定 $r(\delta,p,n)$ 的方法。稍后，Kim 和 Vu [30] 以及 Janson、Oleszkiewicz 和 Ruci´nski [28] 分别独立地得出了这些界。

$$c(\delta)p^2n^2 \;\leqslant\; r(\delta,p,n) \;\leqslant\; C(\delta)p^2n^2\log(1/p)$$

对于某些常数 $c(\delta)$ $C(\delta)$。换言之， $r(\delta,p,n)$ 的值确定到了一个对数误差因子。2010 年前后，Chatterjee [8] 以及 De Marco 和 Kahn [13]分别独立解决了确定 $r(\delta,p,n)$ 正确数量级的问题。他们发现，

$$c(\delta)p^2n^2\log(1/p) \;\leqslant\; r(\delta,p,n) \;\leqslant\; C(\delta)p^2n^2\log(1/p)\,.$$

随后，研究重点转向了对 $r(\delta,p,n)$ 的精确渐近行为。

对于常数 $p \in (0,1)$，Chatterjee 和 Varadhan [11] 证明，求解 $r(\delta,p,n)$ 可归结为求解关于图子的变分问题 [1]。Chatterjee 和 Dembo [10] 将 [11] 的框架扩展到了稀疏情形，前提是 $p \to 0$ 增长相对缓慢，具体来说是 $p \geqslant n^{-1/42}\log n$。通过求解变分问题，Lubetzky 和 Zhao [34]获得了

$$r(\delta,p,n) \;=\; (1+o(1))\min\left\{\frac{\delta^{2/3}}{2},\frac{\delta}{3}\right\}p^2n^2\log(1/p) \tag{1.2}$$

只要 $n^{-1/42}\log n \;\leqslant\; p \;\ll\; 1$。经过 Eldan [16],Cook 和 Dembo [12] 以及 Augeri [1] 的后续改进，我们如今知道，在此情形下，只要满足 $p \gg n^{-1/2}(\log n)^2$，便可应用这一框架。这表明，对所有 $n^{-1/2}(\log n)^2 \ll p \ll 1$，(1.2) 均成立。

最近，Harel、Mousset 和 Samotij [21] 基本上彻底解决了 $G(n,p)$中三角形的大偏差上尾问题。他们的方法结合了早期组合方法中的思想，以及他们提出的熵稳定性新概念。特别是，他们证实只要$p \gg n^{-1/2}$，(1.2) 就成立，并证明

$$r(\delta,p,n) \;=\; (1+o(1))\frac{\delta^{2/3}}{2}p^2n^2\log(1/p)$$

[1]更多关于稠密情形的内容，请参见 Lubetzky 和 Zhao [33] 以及 Chatterjee 的综述 [9] 。

只要 $n^{-1} \ll p \ll n^{-1/2}$，并求出了 $r(\delta, p, n)$的渐近值，剩下的情形是 $p^2 n \to c \in \mathbb{R}$。最后一种情形的表达式更为复杂。当通过添加类似枢纽的结构来“增加” $\delta p^2(n)_3$ 个三角形是“最容易”的方法时，(1.2) 中的$\delta/3$最小；而当通过添加大小适当的团是“最容易”的方法时， $\delta^{2/3}/2$ 最小。在中间情形 $p^2 n \to c$ ，则需要采用混合构造，更多细节见 [21] 。常数平均度的超稀疏随机图这一剩余的“缺失情形”，最近由 Ganguly、Hiesmayr 和 Nam [19]解决。

此外，围绕子图计数中心极限定理这一问题，也有大量研究文献。自 Ruci´nski [40] 确立了子图计数的中心极限定理以来，许多文章相继给出了越来越强的收敛界 [4, 31, 38, 39]。此外，Janson [25] （在早期文章 Janson [24] 以及 Janson 和 Nowicki [27]的基础上）证明了一个函数型中心极限定理，并证明了 $G(n, m)$ 中的子图计数也渐近服从正态分布。

对标准差与均值之间数量级的中等偏大偏差的概率，研究则少得多。D¨oring 和 Eichelsbacher [14] 证明了

$$r(\delta, p, n) \;=\; \frac{-\delta_n^2 p n^2}{36(1-p)} \;+\; o(\delta_n^2 p n^2)\,, \tag{1.3}$$

当 $p^{-1/2} n^{-1} \ll \delta_n \ll p^7$ 时成立（另见 [15]）。而在稠密情形 $p \in (0,1)$ 下，F´eray、M´eliot 和 Nikeghbali [17] 求出了偏差概率的渐近形式（这远强于仅知道速率的渐近形式）：

$$\mathbb{P}\left(N_{\triangle}(G_p) \;>\; (1+\delta) p^3 (n)_3\right) \;= \tag{1.4}$$

$$(1+o(1)) \sqrt{\frac{9(1-p)}{\pi p}} \exp\left(-\frac{\delta_n^2 p n^2}{36(1-p)} \;+\; \frac{(7-8p)\delta_n^3 p n^2}{324(1-p)^2} \;-\; \log(n\delta_n) \right)$$

只要 $n^{-1} \ll \delta_n \ll n^{-1/2}$。

我们可以看出，将这些结果综合起来，它们实际上只涵盖了可能偏差范围中的一小部分。事实上，考虑 $p = n^{\gamma}$ 和 $\delta = n^{\theta}$ 在 $\gamma \in (-1/2, 0)$ 和 $\theta \in (-1, 0)$ 的情形。

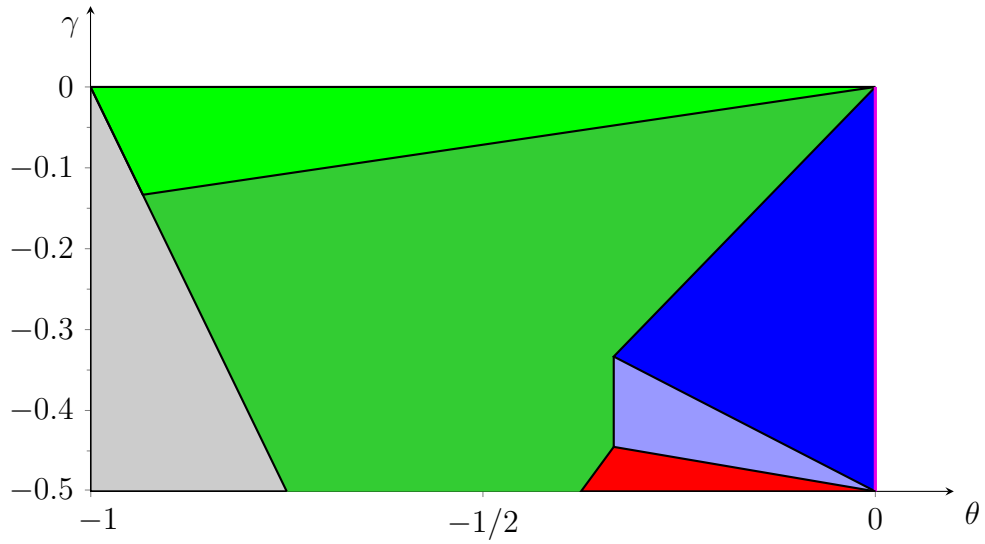


FIGURE 3. D¨oring 和 Eichelsbacher [14] 的结果适用于浅绿色区域。我们将同一界扩展到整个“正态”范围，如图中的绿色区域所示，见定理 1.5。在右侧的三个彩色区域中，会发生局部化（下文将予以描述）。这三种颜色分别对应相应的局部化类型：浅蓝色表示星形图，深蓝色表示枢纽，红色表示团。我们指出，灰色区域对应于小于标准差量级的偏差，根据中心极限定理，其概率为 $\frac{1}{2} + o(1)$。（传统的大偏差结果位于右侧边缘的紫色线段上。）

如果实现偏差的最可能方式涉及随机图中出现一个小型结构，例如星形图、枢纽或团，我们就称发生了局部化。在非局部化区域，偏差更可能是由随机图中出现额外的边（但没有特定结构）造成的。这一概念类似于 Lubetzky 和 Zhao [33] 在稠密情形中考察的副本对称性和对称性破缺。由于我们研究的是稀疏情形，谈论局部化似乎更为自然。Harel、Mousset 和 Samotij [21]表明，在固定 $\delta > 0$ 的情形下，他们的熵稳定性概念蕴含局部化。看起来很有可能对他们的方法稍作调整，便能得到局部化各区间上的正确渐近速率；这些作者最近已针对等差数列的相应结果做到了这一点 [22]。

我们的主要结果之一表明，D¨oring 和 Eichelsbacher [14] 给出的渐近速率 (1.3) 在整个非局部区域内成立。

定理 1.5。设 $n^{-1/2}(\log n)^{1/2} \ll p \ll 1$，并令 δ_n 为一个满足以下条件的序列。

$$p^{-1/2}n^{-1} \ll \delta_n \ll p^{3/4}(\log n)^{3/4}\,,\; n^{-1/3}(\log n)^{2/3} + p\log(1/p)\,.$$

Then

$$r(\delta_n, p, n) \;=\; (1+o(1))\frac{\delta_n^2 pn^2}{36}\,.$$

注 1.6. 关于误差项的真实数量级，详见命题 9.2。

我们的方法不太适用于局部区域。不过，在整个该区域内，我们能够达到 $r(\delta, p, n)$ 的数量级。

定理 1.7. 设 $n^{-1/2}(\log n)^{1/2} \ll p \ll 1$，并令 δ_n 为满足以下条件的序列

$$p^{3/4}(\log n)^{3/4}\,,\; n^{-1/3}(\log n)^{2/3} \;+\; p\log(1/p) \;\leqslant\; \delta_n \;\leqslant 1\,.$$

Then

$$r(\delta_n, p, n) \;=\; \Theta(1)\min\{\delta_n^{2/3}p^2n^2\log n, \delta_n^{1/2}pn^{3/2}\log n \;+\; \delta_n p^2n^2\log(1/p)\}\,.$$

1.3. 樱桃计数的结果。关于樱桃（长度为二的路径）的相应问题要容易一些。我们也不会惊讶于这些关于樱桃的结果能够通过其他方法得到。然而，陈述并证明这些关于樱桃的结果非常有用，因为这为我们的方法提供了一个更初等的应用。我们希望这能帮助读者在更简单的情形下理解该方法的核心思想。

令 $N_\wedge(G)$ 表示图 G 中长度为二的路径的同构副本数（这只是未计重数时计数的两倍）。

我们将再次看到，结果分为不同的区间。这次只有三个。对应的樱桃计数偏差函数为:

$$\begin{aligned}\textsc{normal}_\wedge(b,t) \;&:=\; b^{1/2}tn^{3/2}\,,\\ \textsc{star}_\wedge(b,t) \;&:=\; \frac{b^2}{\ell^2}1_{b<n\ell}\,,\end{aligned}$$

and

$$\textsc{hub}_\wedge(b,t) \;:=\; \frac{bn}{\ell}1_{b\geqslant n\ell}\,.$$

Let

$$M_\wedge(b,t) = \max\{\textsc{normal}_\wedge(b,t), \textsc{star}_\wedge(b,t), \textsc{hub}_\wedge(b,t)\}.$$

并将 DEV∧(*b*, *t*) 定义为使得下式成立的 a 的最小值:

$$\mathbb{P}\left(N_\wedge(G_m) > \mathbb{E}\left[N_\wedge(G_m)\right] + a\right) \;\leqslant\; e^{-b}\,.$$

也就是说，DEV∧(*b*, *t*) 描述的是概率为 e^{-b} 的偏差大小。

我们关于樱桃计数偏差的主要定理，确定了在参数 b 和 t 的几乎整个范围内，DEV∧(*b*, *t*) 的数量级。

定理 1.8. 存在绝对常数 c, C 使得以下结论成立。假设 $2n^{-1}\log\, n \leqslant t \leqslant 1/2$ 且 $3\log\, n \leqslant b \leqslant tn^2\ell$。那么

$$cM_{\wedge}(b,t) \leqslant DEV_{\wedge}(b,t) \leqslant CM_{\wedge}(b,t)\,.$$

我们强调，这一结果是作为次要结果给出的，旨在介绍定理 1.1 上界部分将用到的方法。因此，我们仅对下界给出证明概要，见第 8 节。

1.4. 我们方法的非正式讨论。正如我们所提到的， $G(n,p)$ 情形下的结果将由针对 $G(n,m)$ 模型证明的结果推出。本小节将简要直观地介绍我们在 $G(n,m)$ 中界定三角形计数偏差的方法。

自然，具体方法会因所处区间而异。为明确起见，我们讨论正态区间，即 $\mathrm{NORMAL}(b,t) = b^{1/2}t^{3/2}n^{3/2}$ 是四项中的最大值，因此 $M(b,t) = \mathrm{NORMAL}(b,t)$。在此背景下，我们的目标是证明存在某些常数 $c < C$，使得

$$\mathbb{P}\left(D_{\triangle}(G_m) \geqslant cb^{1/2}t^{3/2}n^{3/2}\right) \geqslant \exp(-b)$$

and

$$\mathbb{P}\left(D_{\triangle}(G_m) \geqslant Cb^{1/2}t^{3/2}n^{3/2}\right) \leqslant \exp(-b)\,.$$

由于 σ 的 $N_{\triangle}(G_m)$标准差的量级为 $t^{3/2}n^{3/2}$，这相当于对中心极限定理的（近似）推广。我们回顾一下，更精确的界已在 [20], 中得到证明，但这些结果似乎难以推广到更稀疏的随机图。

与 [20] 类似，我们的方法基于关于偏差 $D_{\triangle}(G_m)$的鞅表达式，然后利用 Freedman 鞅不等式（我们将在下一节陈述）给出偏差概率的上界和下界。这个表达式（将在第 2.3 节中详细讨论）

$$D_{\triangle}(G_m) = \sum_{i=1}^{m}\left[3\frac{(N-m)_2(m-i)}{(N-i)_3}\,X_{\wedge}(G_i) + \frac{(N-m)_3}{(N-i)_3}\,X_{\triangle}(G_i)\right],$$

将三角形计数偏差表示为增量，这些增量与过程的第 i步中产生的樱桃（$X_{\wedge}(G_i)$）和三角形（$X_{\triangle}(G_i)$）有关。

这种方法面临的挑战包括:

- 理解过程的方差，或者更准确地说，理解其可能的二次变差，以及
- 控制大增量对鞅过程的影响。

如果新边 $e_i = uw$ 使得度 $d(u)$ 或 $d(v)$ 之一与平均值相差甚远，那么项 $X_{\wedge}(G_i)$ 可能会产生很大的增量。项 $X_{\triangle}(G_i)$ 可能

如果新边 $e_i = uw$ 使得共度 $d(u, w)$ 与其期望值相差甚远，那么它可能会产生很大的增量。

因此，我们需要充分理解度、共度及其偏差。我们采用截断和所谓的分治鞅方法。非正式地说，这意味着我们利用对度和共度的理解来控制大增量的影响，同时利用鞅方法控制其余增量。

遗憾的是，正如我们将看到的，这种方法会变得相当复杂。我们希望至少这段非正式的讨论能帮助读者理解后续证明的背景。

1.5. 文章概述。在第 2 节中，我们介绍全文将用到的必要背景和不等式。随后，我们研究 $G(n, m)$ 中度的性质，并在第3节证明若干界。在第4节中，我们证明关于樱桃的定理 1.8（上界部分）。

接下来，我们开始为三角形相关结果做准备。首先，我们在第5节研究 $G(n, m)$ 中的共度。然后，我们在第6节证明所需的偏差上界，并在第7节证明下界；两者合起来即可证明定理 1.1。由于樱桃的下界结果与此类似，我们将在第 8 节讨论它们。

随后，我们将在第9节说明如何将这些结果推广到 $G(n, p)$。第10节将给出一些总结性评论。本文还包含一个附录，其中收录了几项我们认为与正文主题关联不那么直接的证明。

2. Important auxiliary results and inequalities

本节将介绍一些重要的辅助结果和不等式，我们将在全文中使用它们。

我们首先说明，记 N 为 $\binom{n}{2}$，并令 $t := m/N$，它表示随机图 $G_m \sim G(n, m)$ 的密度。我们有时会考虑用于生成 G_m的埃尔德什-雷尼随机图过程。过程 $G_0, G_1, \ldots, G_m$ 可从具有 n 个顶点的空图 G_0 开始生成：每一步都从尚未出现的边中均匀随机地选取一条边并添加进去。等价地，也可以考虑 K_n的边的一个均匀随机排列 $\{e_1, \ldots, e_N\}$，并令图序列 $G_0, \ldots, G_m$ 满足 $E(G_i) = \{e_1, \ldots, e_i\}$。在此背景下，我们用 s 表示图 G_i的密度 $s := i/N$，并注意到$G_i \sim G(n, i)$。

在方便且不太可能造成混淆时，我们会忽略取最接近的整数、向下取整符号等问题。

现在开始陈述所需的辅助结果。

2.1. Chernoff 不等式及相关不等式。我们将陈述适用于二项或超几何随机变量的 Chernoff 不等式，见 [26] 的定理 2.1 和 2.10。注意，最后两个不等式可由 (2.1) 推得。

引理 2.1. 设 X 是一个二项或超几何随机变量，且 $\mu = \mathbb{E}[X]$。那么，对于所有 $a > 0$，我们有

$$\mathbb{P}(X \geqslant \mu + a) \leqslant \exp\left(\frac{-a^2}{2\mu + 2a/3}\right) \tag{2.1}$$

and

$$\mathbb{P}(X \leqslant \mu - a) \leqslant \exp\left(\frac{-a^2}{2\mu}\right). \tag{2.2}$$

对于 $\theta \geqslant e$，我们有

$$\mathbb{P}(X \geqslant \theta\mu) \leqslant \exp\left(-\theta\mu\big(\log(\theta) - 1\big)\right). \tag{2.3}$$

因此，对于 $j \geqslant 3$ 以及任意 $\nu \geqslant \mu$，我们有

$$\mathbb{P}\left(X \geqslant 2^j\nu\right) \leqslant \exp\left(-j2^{j-2}\nu\right). \tag{2.4}$$

有时，考虑由二项分布或超几何分布随机变量之和构成的随机变量也很有用。如果 X 是两个此类随机变量之和，那么由引理 2.1 可知:

$$\mathbb{P}(|X - \mu| \geqslant a) \leqslant 4\exp\left(\frac{-a^2}{8\mu + 2a}\right) \tag{2.5}$$

以及，对于 $j \geqslant 4$，以及任意 $\nu \geqslant \mu$

$$\mathbb{P}\left(X \geqslant 2^j\nu\right) \leqslant 2\exp\left(-j2^{j-3}\nu\right). \tag{2.6}$$

特别是，在计算集合之间边的重数时，也适用这一结论。给定一个图 G，我们可以定义

$$e(U, W) := \sum_{u\in U}\sum_{w\in W} 1_{uw\in E(G)}.$$

注意，令 $X = e(U, W)$，并令 G 服从分布 $G(n, m)$ 或 $G(n, p)$，则 X 是两个二项随机变量或两个超几何随机变量之和，因此不等式 (2.5) 和 (2.6) 成立。

2.2. 鞅偏差不等式。我们从 Hoeffding-Azuma 不等式 [2, 23] 开始。该不等式指出，对于一个超鞅 $(S_i)_{i=0}^m$，其增量为 $(X_i)_{i=1}^m$，有

$$\mathbb{P}(S_m - S_0 > a) \leqslant \exp\left(\frac{-a^2}{2\sum_{i=1}^m c_i^2}\right). \tag{2.7}$$

对所有 $a > 0$，其中 $c_i := \|X_i\|_\infty$。

当增量通常远小于其上确界时，Freedman 的下列不等式给出的界更优。

引理 2.2（Freedman 不等式）。设 $(S_i)_{i=0}^m$ 是关于滤过 $(\mathcal{F}_i)_{i=0}^m$、增量为 $(X_i)_{i=1}^m$ 的超鞅，设 $R \in \mathbb{R}$ 满足 $\max_i |X_i| \leqslant R$ 几乎必然成立，并且对每个 $j \geqslant 1$ 令

$$V(j) := \sum_{i=1}^{j} \mathbb{E}\left[X_i^2 \,\middle|\, \mathcal{F}_{i-1} \right].$$

那么，对于每个 α、$\beta > 0$，我们都有

$$\mathbb{P}\big(S_j - S_0 \geqslant \alpha \quad and \quad V(j) \leqslant \beta \quad for\ some\ j\big) \leqslant \exp\left(\frac{-\alpha^2}{2(\beta + R\alpha)}\right).$$

尤其是，

$$\mathbb{P}\big(S_j - S_0 \geqslant \alpha\big) \leqslant \exp\left(\frac{-\alpha^2}{2(\beta + R\alpha)}\right) + \mathbb{P}\left(V(j) > \beta\right).$$

下面的逆向结果表明，上述界通常接近最优。令 τ_α 为满足 $S_j \geqslant S_0 + \alpha$ 的最小 j（若始终不满足，则取无穷大），并令

$$T_\alpha := V(\tau_\alpha),$$

其中，我们将 $V(\infty)$ 理解为 ∞。上述不等式表明，

$$\mathbb{P}\left(T_\alpha \leqslant \beta\right) \leqslant \exp\left(\frac{-\alpha^2}{2(\beta + R\alpha)}\right).$$

Freedman 逆向不等式 [18]如下。

引理 2.3（Freedman 逆向不等式）。设 $(S_i)_{i=0}^m$ 是一个鞅，其增量为 $(X_i)_{i=1}^m$，相对于滤过 $(\mathcal{F}_i)_{i=0}^m$，设 R 满足 $\max_i |X_i| \leqslant R$ 几乎必然成立，并设 T_α 如上所定义。那么，对任意 α，$\beta > 0$，我们有

$$\mathbb{P}\left(T_\alpha \leqslant \beta\right) \geqslant \frac{1}{2}\exp\left(\frac{-\alpha^2(1+4\delta)}{2\beta}\right),$$

其中 $\delta > 0$ 满足 $\beta/\alpha \geqslant 9R\delta^{-2}$，且 $\alpha^2/\beta \geqslant 16\delta^{-2}\log(64\delta^{-2})$。

我们还将证明 Freedman 不等式的一个推论，并将其应用于 $G(n,m)$ 情形。该推论与 McDiarmid 不等式 [35], 的思路相近；后者考虑满足“Lipschitz”性质的函数的偏差概率。不过，我们研究的是较弱的“Lipschitz”性质，其对应的“Lipschitz 常数”取决于所涉及的边。

令 $\mathcal{G}_{n,m}$ 表示具有 n 个顶点和 m 条边的图族。我们可以考虑图 G、$G' \in \mathcal{G}_{n,m}$ 之间的编辑距离。如果对于某一对 e、$e' \in E(K_n)$，$E(G') = E(G) \setminus \{e\} \cup \{e'\}$，则称两个图 G、G' 相邻（距离为 1）。给定函数 $\psi : E(K_n) \to \mathbb{R}^+$，如果对于每一对相邻的图 G、$G' \in \mathcal{G}_{n,m}$，都有 еиу

$$\left| f(G) - f(G') \right| \leqslant \psi(e) + \psi(e'),$$

其中 $E(G) \triangle E(G') = \{e, e'\}$。

命题 2.4. 设 $G_m \sim G(n,m)$ 且 $t = m/N$。给定 $\psi\colon E(K_n) \to \mathbb{R}^+$ 以及一个 ψ-Lipschitz 函数 $f\colon \mathcal{G}_{n,m} \to \mathbb{R}$，我们有

$$\mathbb{P}\left(f(G_m) - \mathbb{E}\left[f(G_m)\right] \geqslant a\right) \leqslant \exp\left(\frac{-a^2}{24t\|\psi\|^2 + 6a\psi_{\max}}\right)$$

对于所有 $a \geqslant 0$，其中 $\|\psi\|^2 := \sum_{e\in E(K_n)} \psi(e)^2$ 且 $\psi_{\max} := \max_e \psi(e)$。

此外，对于 $\mathbb{P}\left(f(G_m) - \mathbb{E}\left[f(G_m)\right] \leqslant -a\right)$，同样的界也成立。

注 2.5. 我们没有尝试优化常数。

注 2.6。我们将这一结果表述为关于 $G(n,m)$ 的结果，因为这是我们当前所处的研究背景。显然，该结果实际上完全不依赖于图结构。特别地，如果 $f\colon [N]^{(m)} \to \mathbb{R}$，那么对于均匀随机选取的一个 m-元素子集 $G \subseteq [N]$，也可以得到相同的结果。

现在我们给出命题 2.4 的证明。该证明基于 Freedman 不等式的应用，并将依赖于一个以函数 ψ 表示的增量界。

证明。假设 $m \leqslant N/2$ 不失一般性。事实上，若 $m > N/2$，只需注意到 f 也是补图 $G_m^c \sim G_{N-m}$的函数。“此外”部分可通过将不等式应用于 $-f$ 得出。

我们通过考察以下鞅来证明该结果

$$Z_i := \mathbb{E}\left[f(G_m) \,|\, G_i\right] \qquad i = 0, \ldots, m$$

其中 $G_0, \ldots, G_m$ 是 Erd˝os-R´enyi 随机图过程。等价地，令$e_1, \ldots, e_N$为 K_n 的边的一个均匀随机排列，并对每个 i 定义 G_i 为边集为 $\{e_1, \ldots, e_i\}$ 的图。我们需要以下关于该鞅增量的界

$$\left|Z_i - Z_{i-1}\right| \leqslant \psi(e_i) + 2\bar{\psi}\,, \tag{2.8}$$

其中 $\bar{\psi} = N^{-1}\sum_{e\in E(K_n)} \psi(e)$.

下面证明，假设（2.8）成立时，该命题即可得证。在这种情况下，我们可以立即对增量的绝对值作界。

$$\left|Z_i - Z_{i-1}\right| \leqslant \psi(e_i) + 2\bar{\psi} \leqslant 3\psi_{\max}\,,$$

并且我们得到以下界。

$$\begin{aligned}\mathbb{E}\left[\left(Z_i - Z_{i-1}\right)^2 \middle| G_{i-1}\right] &\leqslant \mathbb{E}\left[\left(\psi(e_i) + \frac{2}{N}\sum_{e\in E(K_n)}\psi(e)\right)^2 \middle| G_{i-1}\right] \\ &\leqslant 2\mathbb{E}\left[\psi(e_i)^2|G_{i-1}\right] + \frac{8}{N^2}\left(\sum_{e\in E(K_n)}\psi(e)\right)^2 \\ &\leqslant \frac{4}{N}\sum_{e\in E(K_n)}\psi(e)^2 + \frac{8}{N}\sum_{e\in E(K_n)}\psi(e)^2 \\ &= \frac{12}{N}\|\psi\|_2^2\,.\end{aligned}$$

将上述界对 $i \leqslant m$ 求和，我们得到

$$\sum_{i=1}^{m}\mathbb{E}\left[\left(Z_i - Z_{i-1}\right)^2 \mid G_{i-1}\right] \leqslant \frac{12m}{N}\|\psi\|_2^2 = 12t\|\psi\|_2^2.$$

现在我们将 Freedman 不等式（引理 2.2）应用于 $\alpha = a$、 $\beta = 12t\|\psi\|_2^2$ 和 $R = 3\psi_{\max}$，由此得到

$$\mathbb{P}\left(f(G_m) - \mathbb{E}\left[f(G_m)\right] \geqslant a\right) = \mathbb{P}\left(Z_m - Z_0 \geqslant a\right) \leqslant \exp\left(\frac{-a^2}{24t\|\psi\|_2^2 + 6a\psi_{\max}}\right),$$

如所要求。

剩下的工作是证明（2.8），即，我们必须证明对于每个可能的¯序列（e_1，...，e_i），都有 $|Z_i - Z_{i-1}| \leqslant \psi(e_i) + 2\psi$。条件期望 Z_i 和 Z_{i-1} 可以明确表示为对截至 e_m的边的选择求和，

$$Z_i = \frac{1}{(N-i)_{m-i}}\sum_{e_{i+1},\ldots,e_m} f(G_i \cup \{e_{i+1},\ldots,e_m\})$$

and

$$Z_{i-1} = \frac{1}{(N-i+1)_{m-i+1}}\sum_{e_i',e_{i+1}',\ldots,e_m'} f(G_{i-1} \cup \{e_i', e_{i+1}',\ldots,e_m'\})\,.$$

注意，求和是对与已选边互不相交的互异边序列进行的。我们希望将这些项配对，使图$E(G_i) \cup \{e_{i+1},\ldots,e_m\}$和 $E(G_{i-1}) \cup \{e_i', e_{i+1}',\ldots,e_m'\}$要么相等，要么相邻。显而易见的问题是，这两个求和的项数甚至都不相同。我们在第一个求和中引入一条虚拟边 g，它可以是 $E(K_n) \setminus \{e_1,\ldots,e_{i-1}\}$中的任意一条边。我们有

$$Z_i = \frac{1}{(N-i+1)_{m-i+1}}\sum_{e_{i+1},\ldots,e_m;g} f(G_i \cup \{e_{i+1},\ldots,e_m\})\,.$$

现在，我们可以将两个求和式中的项逐一配对。设 $\mathcal{S}$ 为上述求和式中允许的边序列，$\mathcal{T}$ 为 Z_{i-1} 的求和式中允许的边序列。我们定义一个双射 $\phi\colon \mathcal{S} \to \mathcal{T}$，如下所示：

$$\phi(e_{i+1}, \dots, e_m; g) := \begin{cases} (e_i, e_{i+1}, \dots, e_m) & \text{if } g = e_i \\ (e_{i+1}, \dots, e_{j-1}, e_i, e_{j+1}, \dots, e_m) & \text{if } g \in \{e_{i+1}, \dots, e_m\} \\ (g, e_{i+1}, \dots, e_m) & \text{if } g \notin \{e_i, \dots, e_m\}. \end{cases}$$

注意，在所有情况下，图 $E(G_{i-1}) \cup \{e_i, \dots, e_m\}$ 和 $E(G_{i-1}) \cup \phi(e_{i+1}, \dots, e_m; g)$ 要么相等，要么相邻（对称差为 $\{e_i, g\}$）。现在，由三角不等式，

$$\begin{aligned} &\left|Z_i - Z_{i-1}\right| \\ &\leqslant \frac{1}{(N-i+1)_{m-i+1}} \sum_{e_{i+1},\dots,e_m;g} \left|f(G_{i-1} \cup \{e_i, \dots, e_m\}) - f(G_{i-1} \cup \phi(e_{i+1}, \dots, e_m; g))\right| \\ &\leqslant \frac{1}{(N-i+1)_{m-i+1}} \sum_{e_{i+1},\dots,e_m;g} \psi(e_i) + \psi(g) \\ &\leqslant \psi(e_i) + \frac{1}{N-i+1} \sum_{g \in E(K_n)} \psi(g) \\ &\leqslant \psi(e_i) + 2\bar{\psi}. \end{aligned}$$

证明至此完成。 □

我们还要提及一个相当平凡的界，它只是三角形不等式和并集界的直接推论：

$$\mathbb{P}(X + Y \geqslant \alpha + \beta) \leqslant \mathbb{P}(X \geqslant \alpha) + \mathbb{P}(Y \geqslant \beta) \tag{2.9}$$

对于任意非负随机变量 X、Y 和 α、$\beta \in \mathbb{R}$。

2.3. 樱桃和三角形计数偏差的鞅表示。本小节回顾 [20] 中给出的子图计数偏差的鞅表达式。我们首先一般性地陈述这些结果，然后讨论我们感兴趣的具体情形：樱桃和三角形。

我们首先引入一些记号。对于图 H 和 G，令 $N_H(G)$ 表示 G 中 H 的同构副本数量。也就是说，$N_H(G)$ 表示从 $V(H)$ 到 $V(G)$ 的、保持 H 的边不变的单射数量。这里按重数计数，因此，例如，$N_{K_3}(K_4) = 24$。

如果 H 有 $v(H)$ 个顶点和 $e(H)$ 条边，那么 $N_H(G_m)$ 的期望值为

$$L_H(m) := \mathbb{E}[N_H(G_m)] = \frac{(n)_{v(H)}(m)_{e(H)}}{(N)_{e(H)}}.$$

因此， $N_H(G_m)$ 与其均值的偏差为

$$D_H(G_m) := N_H(G_m) - L_H(m).$$

我们将特别研究 $D_\wedge(G_m)$，即樱桃的偏差，以及$D_\triangle(G_m)$，即三角形的偏差。

设 $G_0, \dots G_m$ 是 Erd˝os-R´enyi 随机图过程的一次实现。稍微滥用一下记号，当我们写 G_m 时，可以假设 G_m 还包含边添加顺序的信息。

如 [20], 中所述，对于每个固定的图 F （尽管我们真正关心的只有 F 为樱桃或三角形的情形），我们可以定义

$$A_F(G_m) := N_F(G_m) - N_F(G_{m-1})$$

它表示添加第 m 条边时所生成的 F 的副本数。我们还可以定义一个“中心化”版本。设

$$X_F(G_m) := A_F(G_m) - \mathbb{E}\left[A_F(G_m) \,|\, G_{m-1}\right] = N_F(G_m) - \mathbb{E}\left[N_F(G_m) \,|\, G_{m-1}\right].$$

$D_H(G_m)$ 的一般鞅表示在 [20] 中给出，为

$$D_H(G_m) = \sum_{i=1}^{m} \sum_{F \subseteq E(H)} \frac{(N-m)_{e(F)}(m-i)_{e(H)-e(F)}}{(N-i)_{e(H)}} X_F(G_i),$$

其中我们对所有满足 $V(F) = V(H)$ 且 $E(F) \subseteq E(H)$ 的 $2^{e(H)}$ 子图 F 求和。由于当 F 没有边或只有一条边时， $X_F(G_i)$ 确定为 0，因此我们得到

$$D_\wedge(G_m) = \sum_{i=1}^{m} \frac{(N-m)_2}{(N-i)_2} X_\wedge(G_i)$$

and

$$D_\triangle(G_m) = \sum_{i=1}^{m} \left[3\frac{(N-m)_2(m-i)}{(N-i)_3} X_\wedge(G_i) + \frac{(N-m)_3}{(N-i)_3} X_\triangle(G_i) \right].$$

我们称这些为鞅表示，因为随机变量 $X_F(G_i)$ 的行为类似于鞅增量，即 $\mathbb{E}\left[X_F(G_i) \,|G_{i-1}\right] = 0$。

2.4. 关于某个求和的界。以下初等不等式将有助于进行一些估计。

命题 2.7. 设 $d \in \mathbb{N}$，设 $r \geqslant d^{1/2}$ 且设 $\beta \geqslant 1$。则，

$$\sum_{x \in \mathbb{Z}^d : \|x\| \geqslant r} \exp\left(-\beta \|x\|^2\right) \leqslant (8\pi)^{d/2} e^{-\beta(r-d^{1/2})^2/2}.$$

为完整起见，证明见附录 A。

3. Degrees in $G(n,m)$

许多论文研究了随机图中的度序列，并且人们已经对度的典型行为有了大量了解。Bollob´as [7] 给出了一篇综述，其中包括 Bollob´as 本人 [5,6] 的研究结果，并提及了其他方法，例如 McKay 和 Wormald [36]的方法。另请参阅 Liebenau 和 Wormald [32] 近期发表的文章及其中的参考文献。

由于我们需要的不仅是关于度的典型行为的结果，还需要了解其在尾部更深处（概率更小的事件）的行为，因此本节陈述我们将要用到的结果。大多数证明都相当标准，因此放在附录中，见附录 B。

在介绍度的结果之后，我们会加入一个简短的小节，说明关于增量 $X_\wedge(G_i)$ 我们可以推导出什么。

我们将在 $G_m \sim G(n,m)$ 中进行研究，不过要指出的是，在 $G(n,p)$ 中也可以轻易推导出类似的结果。设 $d_u(G)$ 是图 G 中顶点 u 的度，并令

$$D_u(G_m) := d_u(G_m) - \frac{2m}{n}$$

b为图 $G_m \sim G(n,m)$ 中 u 的度相对于其均值的偏差 .

每个顶点的期望度为 $2m/n = t(n-1) \approx tn$。我们将给出与度数较大的顶点数量（至少为 $32tn$）以及度数更大的顶点数量（对每个 $j \geqslant 5$，至少为 $2^j tn$ ）相关的界。给出关于 G_m最大度 $\Delta(G_m)$ 的界也很有帮助。这些结果相当标准。我们还将给出与度偏差平方和 $\sum_u D_u(G_m)^2$ 相关的界。

现在我们引入一些记号，以便陈述这些界。令 ℓ_b :=log(b/etn)。设 V_j 表示 $G_m \sim G(n,m)$ 中度数至少为 $2^j tn$ 的顶点集合。

引理 3.1. 假设 $t \geqslant 2n^{-1}\log n$ 且令 $b \geqslant 4tn$。那么，除概率至多为 $\exp(-b)$ 的情形外，我们有

(i) $\Delta(G_m) \leqslant 2b/\ell_b$, *and*

(ii) $|V_j| \leqslant \dfrac{b}{tnj2^{j-6}}$ *for all* $j \geqslant 5$.

关于度偏差平方和，我们发现尾部不同位置存在三种不同的行为。令 $\kappa(b,t)$ 为如下定义的函数:

$$\kappa(b,t) := \begin{cases} tn^2 & 1 \leqslant b < t^{1/2}n\ell \\ b^2/\ell^2 & t^{1/2}n\ell \leqslant b < n\ell \\ bn/\ell & n\ell \leqslant b \leqslant tn^2\ell\,. \end{cases}$$

这三个范围分别对应度偏差平方和较大的三种不同“成因”。第一个值对应于约有 n 个顶点的贡献，这些顶点的度偏差为标准差，约为 $t^{1/2}n^{1/2}$。第二个值和第三个值分别与星形子图和枢纽造成的度偏差有关。我们还定义

$$\kappa^+(b,t) := \begin{cases} b^2/\ell_b^2 & 32tn \leqslant b < n\ell \\ bn/\ell & n\ell \leqslant b \leqslant tn^2\ell\,. \end{cases}$$

现在我们可以给出偏差平方和的界。

命题 3.2. 存在一个绝对常数 $C > 0$，使得以下结论成立。假设 $t \geqslant 2n^{-1}\log n$ 且 $b \geqslant 32tn$。除概率至多为 $\exp(-b)$ 的情况外，我们有

$$\sum_u D_u(G_i)^2 \leqslant C\kappa(b,t)$$

对于所有步骤 $i \leqslant m$，以及

$$\sum_{u\,:\,d_u(G_m)\geqslant 32tn} d_u(G_m)^2 \leqslant C\kappa^+(b,t)\,.$$

下面的引理界定了度并不太大的顶点所作的贡献，涵盖了命题 3.2 证明中最有趣的部分。正因如此，我们在此给出引理 3.3 的证明，而命题 3.2 证明的其余部分则见附录 B。

引理 3.3. 存在绝对常数 $C > 0$，使得以下结论成立。假设 $t \geqslant 2n^{-1}\log n$ 且 $b \geqslant n$。除去概率至多为 $\exp(-b)$ 的情况，我们有

$$\sum_{u:d_u(G_i)\leqslant 32tn} D_u(G_i)^2 \leqslant Cbtn$$

对于所有步骤 $i \leqslant m$。

我们希望使用鞅方法来界定以下和$\sum_{u:\ d_u(G_i)\leqslant 32tn} D_u(G_i)^2$ 很大的概率。可能使分析复杂化的一点是这个表达式的非线性。下面的技巧可以将其“线性化”。如果度偏差平方和很大，那么我们可以用一个向量 σ 来编码它，使该向量模拟度偏差，从而 $\sum_{u\in V(G_m)}\sigma_u D_u(G_i)$也很大。由于这个函数是线性的，并且具有某些 Lipschitz 性质，我们可以利用命题 2.4 界定它很大的概率。将这一想法与并集界结合起来，即可得到所需结果。

σ 的作用是将每个顶点的（重新归一化的）度偏差编码到一个二倍因子的范围内。因此，读者可以将 σ 视为重新归一化的度偏差的二进制舍入。那些小到无关紧要的值则设为 0。

引理 3.3 的证明。给定一个向量 $\sigma \in \mathbb{Z}^{V(G)}$，其各项均为 0 或$\pm 2^j$，其中 $j = 0$取值为 $\ldots, \lfloor \frac{1}{2}\log_2 tn \rfloor$，我们对每个 $i \leqslant m$定义事件 $E_{\sigma,\, i}$，使得

$$\sum_{u \in V(G_i)} \sigma_u D_u(G_i) \geqslant 16 \|\sigma\|^2 t^{1/2} n^{1/2} . \tag{3.1}$$

令 S 为满足 $\|\sigma\|^2 \geqslant 16b$ 的这类向量 σ 的集合。我们提出两个断言：

断言 1：事件 $\sum_{u:\ d_u(G_i) \leqslant 32tn} D_u(G_i)^2 > 2^{20} btn$ 包含于并集 $\bigcup_{\sigma \in S} E_{\sigma,\, i}$ 中。

断言 2：对于所有 $\sigma \in S$ 和 $i \leqslant m$，$\mathbb{P}(E_{\sigma,\, i}) \leqslant \exp(-\|\sigma\|^2)$。

注意到，该引理可由这两个断言推出。由断言 1、并集界以及断言 2 可知，存在某个 $i \leqslant m$ 使得 $\sum_{u:\ d_u(G_i) \leqslant 32tn} D_u(G_i)^2 > 2^{20} btn$ 发生的概率至多为 $\sum_{i=1}^{m} \sum_{\sigma \in S} \mathbb{P}(E_{\sigma,\, i}) \leqslant m \sum_{\sigma \in S} \exp(-\|\sigma\|^2)$。现在，由推论 2.7，取 $r = 4b^{1/2}$，我们有

$$\begin{aligned} m \sum_{\sigma \in S} \exp(-\|\sigma\|^2) &\leqslant n^2 (8\pi)^{n/2} \exp\left(\frac{-(4b^{1/2} - n^{1/2})^2}{2} \right) \\ &\leqslant n^2 \exp(2n - 9b/2) \\ &\leqslant \exp(-b) , \end{aligned}$$

如所要求的。

断言 I 的证明：如果图中发生事件 $\sum_{u:\ d_u(G_i) \leqslant 32tn} D_u(G_i)^2 > 2^{20} btn$，那么我们将证明，对于下面给出的 σ 的取值，E_σ 成立。令 U 表示满足 $|D_u(G_i)| \in [32t^{1/2} n^{1/2}, 32tn]$ 的顶点集合。偏差小于 $32t^{1/2} n^{1/2}$ 的顶点所产生的贡献至多为 $2^{10} tn^2 \leqslant 2^{10} btn$，因此

$$\sum_{u \in U} D_u(G_i)^2 > 2^{19} btn . \tag{3.2}$$

定义 σ 为 $\sigma_u = 0$，若 $u \notin U$；对于 $j \geqslant 0$，我们设定

$$\sigma_u := \begin{cases} 2^j & 2^{j+5} t^{1/2} n^{1/2} \leqslant D_u(G_i) < \min\{2^{j+6} t^{1/2} n^{1/2}, 32tn\} \\ -2^j & 2^{j+5} t^{1/2} n^{1/2} \leqslant -D_u(G_i) < \min\{2^{j+6} t^{1/2} n^{1/2}, 32tn\} . \end{cases}$$

根据 σ 的定义可知，

$$\frac{D_u(G_i)^2}{2^{12} tn} \leqslant \sigma_u^2 \leqslant \frac{D_u(G_i)^2}{2^{10} tn}$$

对所有 $u \in U$ 均成立。因此，根据 (3.2)，我们有

$$16b \leqslant 2^7 b \leqslant \|\sigma\|^2 \leqslant \frac{1}{2^{10} tn} \sum_{u \in U} D_u(G_i)^2 .$$

特别是，下界证实了 $\sigma \in S$。我们还注意到，对于所有 $u \in U$， $\sigma_u D_u(G_i) \geqslant D_u(G_i)^2/2^6 t^{1/2} n^{1/2}$，因此

$$\begin{aligned}\sum_{u \in V(G_i)} \sigma_u D_u(G_i) &\geqslant \frac{1}{2^6 t^{1/2} n^{1/2}} \sum_{u \in U} D_u(G_i)^2 \\ &\geqslant 16 t^{1/2} n^{1/2} \|\sigma\|^2 .\end{aligned}$$

这表明 E_σ 发生，从而完成了断言 1 的证明。

断言 2 的证明：令我们固定 $\sigma \in S$。设

$$f_\sigma(G_i) := \sum_{u \in V(G_i)} \sigma_u D_u(G_i) .$$

因此，事件 E_σ 就是函数 $f_\sigma(G_i)$ 至少为 $16 t^{1/2} n^{1/2} \|\sigma\|^2$ 的事件。我们利用命题 2.4 中的不等式来界定此概率。用该命题的术语来说，对于函数 $\psi(uw) = |\sigma_u| + |\sigma_w|$，我们的函数 f_σ 是 ψ-Lipschitz 的。注意， $\mathbb{E}[f_\sigma(G_i)] = 0$，并且 ψ 满足 $\|\psi\|^2 \leqslant 2n \|\sigma\|^2$ 和 $\psi_{\max} \leqslant 2\sigma_{\max} \leqslant 2t^{1/2} n^{1/2}$。应用命题 2.4，我们得到

$$\begin{aligned}\mathbb{P}(E_\sigma) &= \mathbb{P}\left(f_\sigma(G_i) \geqslant 16 \|\sigma\|^2 t^{1/2} n^{1/2}\right) \\ &\leqslant \exp\left(\frac{-256 \|\sigma\|^4 tn}{48tn \|\sigma\|^2 + 192tn \|\sigma\|^2}\right) \\ &\leqslant \exp(-\|\sigma\|^2) .\end{aligned}$$

这就完成了证明。 □

3.1. $X_\wedge(G_i)$ 的增量。现在，我们根据关于度偏差平方和的结果，推导出以下增量界。

引理 3.4. 存在一个绝对常数 C，使得下述结论成立。设 $2n^{-1} \log n \leqslant t \leqslant 1/2$，且 $b \geqslant 32tn$，并且 $i \leqslant m$。那么，除概率至多为 $\exp(-b)$ 的情况外，我们有

$$\mathbb{E}\left[X_\wedge(G_i)^2 \middle| G_{i-1}\right] \leqslant \frac{C\kappa(b,t)}{n} .$$

证明。回忆一下， $X_\wedge(G_i)$ 定义为 $X_\wedge(G_i) := A_\wedge(G_i) - \mathbb{E}[A_\wedge(G_i)|G_{i-1}]$，它表示 $A_\wedge(G_i)$ 与其在给定 G_{i-1} 时的期望值之差；其中， $A_\wedge(G_i)$ 是添加第 i 条边所产生的长度为 2 的路径（同构副本）的数量。因此，

$$\mathbb{E}\left[X_\wedge(G_i)^2 \middle| G_{i-1}\right] = \mathrm{Var}(X_\wedge(G_i)|G_{i-1}) = \mathrm{Var}(A_\wedge(G_i)|G_{i-1}) .$$

现在利用这一点：对于任意常数 c，我们有 $\mathrm{Var}(X) \leqslant \mathbb{E}\,[(X-c)^2]$,，因此

$$\mathbb{E}\left[X_\wedge(G_i)^2 \middle| G_{i-1}\right] \leqslant \mathbb{E}\left[\left(A_\wedge(G_i) - \frac{8(i-1)}{n}\right)^2 \middle| G_{i-1}\right].$$

通过计算所产生的樱桃的同构副本数量，我们看到 $A_\wedge(G_i) = 2d_u(G_{i-1}) + 2d_w(G_{i-1})$，其中 uw 是 G_i中纳入的第 i 条边。现在，由于 uw 是从 $E(K_n) \setminus E(G_{i-1})$中均匀选取的一对，我们有

$$\begin{aligned}
\mathbb{E}[X_\wedge(G_i)^2 | G_{i-1}] &\leqslant \frac{1}{N-i+1} \sum_{uw \notin E(G_{i-1})} \left(2d_u(G_{i-1}) + 2d_w(G_{i-1}) - \frac{8(i-1)}{n}\right)^2 \\
&\leqslant \frac{8}{N-i+1} \sum_{uw \notin E(G_{i-1})} \left(d_u(G_{i-1}) - \frac{2(i-1)}{n}\right)^2 + \left(d_w(G_{i-1}) - \frac{2(i-1)}{n}\right)^2 \\
&\leqslant \frac{32}{n^2} \sum_{uw \notin E(G_{i-1})} D_u(G_{i-1})^2 + D_w(G_{i-1})^2 \\
&\leqslant \frac{32}{n} \sum_u D_u(G_{i-1})^2 .
\end{aligned}$$

所需的不等式现在可由命题 3.2 立即推出。 □

4. Upper bound on cherry count deviation in $G(n,m)$

本节证明定理 1.8 的上界部分。这对应于下文陈述的命题 4.1。我们回顾一下,

$$M_\wedge(b,t) := \max\{\mathrm{NORMAL}_\wedge(b,t), \mathrm{STAR}_\wedge(b,t), \mathrm{HUB}_\wedge(b,t)\}$$

其中 $\mathrm{NORMAL}_\wedge(b,t) := b^{1/2}tn^{3/2}$，$\mathrm{STAR}_\wedge(b,t) := b^2 1_{b<n\ell}/\ell^2$，以及 $\mathrm{HUB}(b,t) := bn1_{b\geqslant n\ell}/\ell$。以下陈述对应于定理 1.8 的上界部分。

命题 4.1。存在一个绝对常数 C，使得下述结论成立。对于所有 $2n^{-1}\log n \leqslant t \leqslant 1/2$ 和 $3\log n \leqslant b \leqslant tn^2\ell$，我们有

$$\mathbb{P}\left(D_\wedge(G_m) \geqslant CM_\wedge(b,t)\right) \leqslant \exp(-b).$$

T该证明是一个相对直接的分治鞅论证。 t.

在给出证明之前，先对 $M_\wedge(b,t)$ 作一些观察是有益的。容易验证，对于 t 的每一种取值，当 $b \leqslant t^{2/3}n\ell^{4/3}$ 时，正态项 $\mathrm{NORMAL}_\wedge(b,t)$ 最大；当 $t^{2/3}n\ell^{4/3} < b \leqslant n\ell$ 时，星项 $\mathrm{STAR}_\wedge(b,t)$ 最大；而当 $b > n\ell$ 时，枢纽项 $\mathrm{HUB}(b,t) := bn1_{b\geqslant n\ell}/\ell$ 最大。还可以容易地验证,

$$M_\wedge(b,t) \geqslant btn \tag{4.1}$$

对所有 $b \geqslant 1$，且

$$M_\wedge(b,t)^2 \geqslant btn\kappa(b,t) \tag{4.2}$$

对所有 $b \geqslant 1$，其中 $\kappa(b,t)$ 的定义见第3节。为证明 (4.2)，注意到对于 $1 \leqslant b \leqslant t^{1/2}n\ell$，我们有 $M_\wedge(b,t)^2 \geqslant \mathrm{NORMAL}_\wedge(b,t)^2 = bt^2n^3 = btn\kappa(b,t)$；而对于 $b \geqslant t^{1/2}n\ell$，我们有 $M_\wedge(b,t) = \kappa(b,t)$，因此 (4.2) 可由 (4.1) 得出。

另外，注意到不等式

$$M_\wedge(b,t) \geqslant \frac{1}{9}\kappa^+(b,t) \tag{4.3}$$

对所有 $b \geqslant 32tn$ 均成立。我们考虑三个范围。对于 $32tn \leqslant b \leqslant t^{2/3}n\ell_b^{4/3}$，这由以下事实可知:

$$M_\wedge(b,t) \geqslant \mathrm{NORMAL}_\wedge(b,t) = b^{1/2}tn^{3/2} \geqslant \frac{b^2}{\ell_b^2} = \kappa^+(b,t)$$

在此区间内。容易看出，对于 $b \geqslant 32tn$，有 $\ell_b^{4/3} \geqslant e$，因此，对于 $b > \max\{32tn, t^{2/3}n\ell_b^{4/3}\}$，有 $\ell_b \geqslant \ell/3$。所以，在 $\max\{32tn, t^{2/3}n\ell_b^{4/3}\} \leqslant b \leqslant n\ell$ 的范围内，我们有

$$M_\wedge(b,t) \geqslant \mathrm{STAR}_\wedge(b,t) = b^2/\ell^2 \geqslant \frac{b^2}{9\ell_b^2} = \frac{1}{9}\kappa^+(b,t)$$

在此区间内。最后，对于 $b > n\ell$，有 $M_\wedge(b,t) = $ 枢纽$_\wedge(b,t) = \kappa^+(b,t)$。

我们指出，对于星形图和枢纽区域，实际上可以通过将樱桃计数表示为度偏差平方和，从命题 3.2 更直接地推导出这些结果。不过，我们更愿意给出一个适用于各个区间的统一证明。这样做的一个原因是，该证明概述了我们稍后在三角形情形中还会用到的证明技巧。

我们回顾一下，$D_\wedge(G_m)$ 的鞅表达式为

$$D_\wedge(G_m) = \sum_{i=1}^{m} \frac{(N-m)_2}{(N-i)_2} X_\wedge(G_i)\,.$$

为避免某个增量产生过大的影响，我们考虑对增量进行截断。令

$$X'_\wedge(G_i) := X_\wedge(G_i)1_{X_\wedge(G_i)\leqslant 128tn}$$

and

$$Z_\wedge(G_i) := X_\wedge(G_i)1_{X_\wedge(G_i)>128tn}\,.$$

由此可得

$$D_\wedge(G_m) = D'_\wedge(G_m) + N^*_\wedge(G_m)$$

其中

$$D'_\wedge(G_m) := \sum_{i=1}^{m} \frac{(N-m)_2}{(N-i)_2} X'_\wedge(G_i) \tag{4.4}$$

and

$$N^*_\wedge(G_m) := \sum_{i=1}^{m} \frac{(N-m)_2}{(N-i)_2} Z_\wedge(G_i)\,. \tag{4.5}$$

因此，要证明命题 4.1，只需证明存在绝对常数 C_1 和 C_2，使得

$$\mathbb{P}\left(D'_\wedge(G_m) \geqslant C_1 M_\wedge(b,t)\right) \leqslant \exp(-2b) \tag{4.6}$$

and

$$\mathbb{P}\left(N^*_\wedge(G_m) \geqslant C_2 M_\wedge(b,t)\right) \leqslant \exp(-2b) \tag{4.7}$$

因此我们可以取 $C = C_1 + C_2$，所需的界便由三角不等式和并集界推出，正如 (2.9) 中所示。

4.1. 对 $D'_\wedge(G_m) \geqslant C_1 M_\wedge(b,t)$ 的概率进行界定。我们将使用 Freedman 不等式（引理 2.2）来证明 (4.6)。

由于 $X'_\wedge(G_i) \leqslant X_\wedge(G_i)$，(4.4) 中求和的部分和序列构成一个超鞅，其最终值为 $D'_\wedge(G_m)$。我们首先考察增量 Y'_i 的行为，其定义如下

$$Y'_i := \left((N-m)_2/(N-i)_2\right) X'_\wedge(G_i)\,.$$

我们立即注意到，由于系数至多为 1，因此有确定性上界 $|Y'_i| \leqslant 128tn$。

我们还可以对该过程的二次变差取界。注意，$(Y'_i)^2 \leqslant (X'_\wedge(G_i))^2 \leqslant X_\wedge(G_i)^2$。因此，将引理 3.4 应用于“$b$”为 $\max\{4b, 32tn\}$，可得一个常数 C_3，使得对每个 $i \leqslant m$，概率至多为 $\exp(-4b)$，即存在 հնարավորություն

$$\mathbb{E}\left[(Y'_i)^2 | G_{i-1}\right] > C_3 n^{-1} \kappa(b,t)\,.$$

因此，除了概率至多为 $\exp(-3b)$ 的情况外，我们有

$$\sum_{i=1}^{m} \mathbb{E}\left[(Y'_i)^2 | G_{i-1}\right] \leqslant C_3 \kappa(b,t) tn\,.$$

设 $\beta = C_3\kappa(b,t)tn$。现在我们应用 Freedman 不等式，即引理 2.2。由此得到的是

$$\mathbb{P}\left(D'_\wedge(G_m) \geqslant C_1 M_\wedge(b,t)\right) \leqslant \exp\left(\frac{-C_1^2 M_\wedge(b,t)^2}{2\beta + 256 C_1 M_\wedge(b,t) tn}\right) + \exp(-3b)\,.$$

由 (4.1) 和 (4.2) 立即可知，只要将 C_1 取得足够大，第一项至多为 $\exp(-3b)$。由于 $2\exp(-3b) \leqslant \exp(-2b)$，证明至此完成。

4.2. 界定 $N^*_\wedge(G_m) \geqslant C_2 M_\wedge(b,t)$ 的概率。为了证明 (4.7)，我们将利用第3节中证明的关于度的行为的结果来界定 $N^*_\wedge(G_m)$。我们首先建立 $N^*_\wedge(G_m)$ 与度序列之间的联系。

由 $X_\wedge(G_i)$ 的定义可知

$$X_\wedge(G_i) \leqslant A_\wedge(G_i) \leqslant 2\sum_{v \in e_i} d_v(G_{i-1})\,.$$

特别地，若 $X_\wedge(G_i) > 128tn$，则度较大的顶点的度至少为$32tn$。因此

$$N^*_\wedge(G_m) \;=\; \sum_{i=1}^{m} X_\wedge(G_i) 1_{X_\wedge(G_i)>128tn} \;\leqslant\; 4\sum_{i=1}^{m}\sum_{v\in e_i} d_v(G_m) 1_{d_v(G_m)>32tn}\,.$$

注意，在最后一个不等式中，我们用到了 $d_v(G_{i-1}) \leqslant d_v(G_m)$。现在，通过双重计数可得

$$\begin{aligned} N^*_\wedge(G_m) \;&\leqslant\; 4\sum_{v\in V} d_v(G_m) 1_{d_v(G_m)>32tn} \sum_{e\in E(G_m)} 1_{v\in e} \\ &= 4\sum_{v\in V} d_v(G_m)^2 1_{d_v(G_m)>32tn}\,. \end{aligned}$$

如果 $\Delta(G_m) \leqslant 32tn$，右侧的量自动为 0。特别是，如果 $b \leqslant 16tn$，那么将引理 3.1 中的“b”取为 $32tn$，可得

$$\mathbb{P}\left(\Delta(G_m) > 32tn\right) \;\leqslant\; \exp(-32tn) \;\leqslant\; \exp(-2b)\,.$$

因此，在这种情况下，$N^*_\wedge(G_m)$ 非零的概率至多为 $\exp(-2b)$，于是 (4.7) 立即成立。

对于 $b \geqslant 16tn$，应用命题 3.2 可得，除概率至多为 $\exp(-2b)$ 的情况外，有

$$N^*_\wedge(G_m) \;\leqslant\; C_2\kappa^+(b,t) \;\leqslant\; 9C_2 M_\wedge(b,t)\,,$$

最后一个不等式用到了 (4.3)。至此，(4.7) 的证明完成。

5. G(n, m) 中的余邻数

上一节中关于樱桃的证明依赖于对度序列的一些控制。随机变量 $X_\triangle(G_i)$ 的二次变差是共度序列的函数。因此，要将这一方法推广到三角形计数，我们还需要证明与共度序列的行为有关的界。事实证明，这是一项艰巨得多的挑战，因为我们需要研究共度数较大的顶点对集合的结构。

证明这些结果之后，我们将在第 5.3 节说明，根据这些关于共度行为的界，可以对相关增量 $X_\triangle(G_i)$ 的二次变差作出哪些推论。

对于每一对不同顶点 $u, w \in V(G_m)$（等价地，对于每个 $uw \in E(K_n)$），我们用 $d_{uw}(G_m)$ 表示它们的共度，即 $d_{uw}(G_m) = |N_{G_m}(u) \cap N_{G_m}(w)|$，也就是图 G_m 中 u 和 w 的公共邻点数。我们还定义

$$D_{uw}(G_m) \;=\; d_{uw}(G_m) \;-\; \frac{m(m-1)(n-2)}{N(N-1)}\,,$$

即共度相对于其在 $G_m \sim G(n,m)$ 中的期望值的偏差。我们将考察共度偏差的两个范围

(i) 从 $\Theta(tn^{1/2})$ 到 $\Theta(t^2n)$,
以及 (ii) 大于 $\Theta(t^2n)$。
第一个范围考虑从共度标准差的量级到均值量级的偏差，而第二个范围考虑真正较大的共度偏差。

一个观察是，度较大的顶点 u 很可能与几乎所有顶点都有较大的共度。因此，本节聚焦于共度较大的点对 uw，而这种较大共度无法直接用其中一个顶点的度来解释。与上文的 (i) 和 (ii) 相对应，

(i) 对于 $k \in K_1 := \{10, \dots, \lceil \log_2(tn^{1/2}) + 10 \rceil\}$，我们定义

$$F_k(G) := \{uw : |D_{uw}(G)| \in [2^k tn^{1/2}, 2^{k+1}tn^{1/2}) \quad |D_u(G)|, |D_w(G)| \leqslant 2^{k-5}n^{1/2}\}$$

并令 $f_k(G) := |F_k(G)|$。

(ii) 对于 $k \in K_2 := \{10, \dots, \lfloor 2\log_2(1/t) \rfloor\}$，我们定义

$$H_k(G) := \{uw : d_{uw}(G) \in [2^k t^2 n, 2^{k+1}t^2 n)\,,\, d_u(G), d_w(G) \leqslant 2^{k-5}tn\}$$

$$\text{and set } h_k(G) := |H_k(G)|.$$

注意，区间 K_1 和 K_2 的选取方式使得从 $2^{10}tn^{1/2}$ 到 n 的所有可能共度的二进区间都被覆盖。细心的读者会注意到，我们在定义 $H_k(G)$时使用的是共度 $d_{uw}(G)$，而不是共度偏差 $D_{uw}(G)$。这样做是因为讨论共度更容易，而且在应用中，只要 $d_{uw}(G_i) \geqslant t^2n$，我们就会使用 $|D_{uw}(G_i)| \leqslant d_{uw}(G_i)$。

我们将证明以下界。

命题 5.1. 存在一个绝对常数 C，使得以下结论成立。设 $t \geqslant Cn^{-1/2}(\log n)^{1/2}$ 且 $b \geqslant n$。那么，除概率至多为 $\exp(-b)$ 的情况外，我们有

$$f_k(G_i) \leqslant \frac{Cb^2}{2^{4k}} \qquad \textit{for all } k \in K_1 \textit{ and } i \leqslant m\,.$$

对于 $b \geqslant 3\log\, n$，除概率至多为 $\exp(-b)$ 的情况外，我们有

$$h_k(G_i) \leqslant \frac{Cb^2}{k^2 2^{2k} t^4 n^2} \qquad \textit{for all } k \in K_2 \textit{ and } i \leqslant m\,.$$

为了应用于控制三角形计数，尤其是相关过程的二次变差，我们将使用以下推论。

推论 5.2. 存在一个绝对常数 C，使得以下结论成立。假设 $t \geqslant Cn^{-1/2}(\log n)^{1/2}$ 且 $b \geqslant n$。那么，除概率为 at

至多 $\exp(-b)$，我们有

$$\sum_{uw} D_{uw}(G_i)^2 \leqslant C\max\{bt^2n^2, b^2\}$$

对所有 $i \leqslant m$ 均成立。

证明。我们可以固定 $i \leqslant m$，并证明该结果成立的失败概率至多为 $4\exp(-2b)$。然后由并集界可得该推论，因为 $4m\exp(-2b) \leqslant 4n^2\exp(-2b) \leqslant \exp(-b)$。

我们将考察四类对，它们对和式 $\sum_{uw} D_{uw}(G_i)^2$ 有贡献。

首先，满足 $|D_{uw}(G_i)| \leqslant 2^{10}tn^{1/2}$ 的点对总贡献至多为 $2^{20}t^2n^3 \leqslant 2^{20}bt^2n^2$，因此可以放心忽略。

其次，满足 $|D_{uw}(G_i)| \leqslant 64t\max\{|D_u(G_i)|, |D_w(G_i)|\}$ 的点对，其贡献可由度偏差平方和的界控制。事实上，根据命题 3.2，除失败概率至多为 $\exp(-32b)$ 的情形外，这些项的总贡献至多为

$$2^{12}t^2n\sum_u D_u(G_i)^2 \leqslant C_1t^2n\kappa(32b,t) \leqslant \frac{C_1'bt^2n^2}{\ell}$$

对某些常数 C_1 和 C_1'。

所有其余点对都出现在 $\bigcup_{k\in K_1} F_k(G_i)$ 或 $\bigcup_{k\in K_2} H_k(G_i)$ 中。除概率至多为 $\exp(-2b)$ 的情形外，我们有

$$f_k(G_i) \leqslant \frac{C_2b^2}{2^{4k}} \qquad \text{for all } k \in K_1, \text{ and}$$

$$h_k(G_i) \leqslant \frac{C_2b^2}{k^22^{2k}t^4n^2} \qquad \text{for all } k \in K_2\,,$$

对某个常数 C_2。下文将假设这些界成立。

我们先考虑 $\bigcup_{k\in K_1} F_k(G_i)$ 中的点对。若 $uw \in F_k(G_i)$，则 $|D_{uw}(G_i)| \leqslant 2^{k+1}tn^{1/2}$，因此我们可以用 $2^{2k+2}t^2nf_k(G_i)$ 界定 $F_k(G_i)$ 中点对的总贡献。

我们分两种情况讨论。若 $2^{2k} \leqslant b/n$，则使用平凡界 $f_k(G_i) \leqslant n^2$，因此 $F_k(G_i)$ 中点对的总贡献至多为

$$2^{2k+2}t^2nf_k(G_i) \leqslant 2^{2k+2}t^2n^3\,.$$

将这些贡献相加，至多为 $8bt^2n^2$。对于 k，当 $2^{2k} > b/n$ 时，我们使用上述关于 $f_k(G_i)$ 的界，从而得到 $F_k(G_i)$ 中点对的总贡献至多为

$$2^{2k+2}t^2nf_k(G_i) \leqslant \frac{4C_2b^2t^2n}{2^{2k}}\,.$$

将这些贡献相加，至多得到 $8C_2bt^2n^2$。因此

$$\sum_{k\in K_1}\sum_{uw\in F_k(G_i)} D_{uw}(G_i)^2 \leqslant 8(C_2+1)bt^2n^2\,.$$

现在考虑 $\bigcup_{k\in K_2} H_k(G_i)$ 中的点对。如果 $uw \in H_k(G_i)$，那么 $D_{uw}(G_i)$ 6d_{uw} $(G_i) \leqslant 2^{k+1}t^2n$ ，因此我们可以将 $H_k(G_i)$ 中点对的总贡献界定为 $2^{2k+2}t^4n^2h_k(G_i)$。对 $k \in K_2$ 上的这些贡献求和，至多得到

$$\sum_{k\in K_2} 2^{2k+2}t^4n^2h_k(G_i) \leqslant \sum_{k\geqslant 10}\frac{4C_2b^2}{k^2} \leqslant C_2b^2\,.$$

S将所有类型的贡献全部相加，可得该求和至多为

$$\left(2^{10}+C_1+9C_2+8\right)\max\{bt^2n^2, b^2\}\,,$$

如所要求的。 □

在本节余下部分，我们证明命题 5.1。对于 $f_k(G_i)$ 和 $h_k(G_i)$，证明略有不同，因此我们分别为它们安排一个小节。事实上，最好从 $h_k(G_i)$ 开始。

5.1. 控制 $h_k(G_i)$。现在我们将证明命题 5.1 中关于 $h_k(G_i)$ 的部分。

我们注意到，当 $k2^k \leqslant b/t^2n^2$ 时，由于 $h_k(G) \leqslant n^2$，该界是平凡的。因此，让我们固定 $i \leqslant m$ 和 $k \in K_2$ ，使得 $k2^k > b/t^2n^2$。

我们可以将 $H_k(G)$ 视为与图 G 相关联的辅助图。与其直接处理 $H_k(G)$，不如根据 u、w 在对 uw 中的度，对 $H_k(G)$ 进行划分。我们定义

$$H_{k,0}(G) := \{uw : d_{uw}(G) \in [2^kt^2n, 2^{k+1}t^2n)\,,\ \max\{d_u(G), d_w(G)\} \leqslant 2^{k/2}tn\}$$

a以及集合 $h_{k,0}(G) := |H_{k,0}(G)|$。对于 $k/2 < j \leqslant \min\{k-6, \log_2(1/t)\}$，我们定义 e

$$H_{k,j}(G) := \{uw:\ d_{uw}(G) \in [2^kt^2n, 2^{k+1}t^2n)\,,\ \max\{d_u(G)\ d_w(G)\} \in [2^jtn, 2^{j+1}tn)\ \}$$

并令 $h_{k,j}(G) := |H_{k,j}(G)|$。

令 $J = \{0\} \cup \{j :\ k/2 < j \leqslant \min\{k-6, \log_2(1/t)\}\}$ 为指标集合。由于 $H_k(G)$ 中的每一对都出现在某个 $H_{k,j}(G)$ 中，且满足 $j \in J$，因此我们有

$$h_k(G) = \sum_{j\in J} h_{k,j}(G)\,. \tag{5.1}$$

下面的引理给出了辅助图$H_{k,\ 0}(G)$ 和 $H_{k,\ j}(G)$ 中某些结构大小的界。事实上，该引理将在图$G \sim G(n,p)$ 的背景下表述，其中 $p \in (0,t)$。这样可以使证明稍微简洁一些，因为我们避免了与条件化相关的一些问题。

引理 5.3. 假设 $t \geqslant n^{-1/2}(\log n)^{1/2}$，$p \in (0,t)$，且 $b \geqslant 3\log n$。令$10 \leqslant k \leqslant 2\log_2(1/t)$，并令 $k/2 < j \leqslant \min\{k-6, \log_2(1/t)\}$，再令 $G \sim G(n,p)$。其概率至少为 $1-\exp(-b)$：

(i) $H_{k,0}(G)$ 不包含度至少为 $32b/k2^k t^2 n$的星形图；(ii) $H_{k,j}(G)$ 不包含度至少为 $32b/(k-j)2^k t^2 n$的星形图；(iii) $H_{k,0}(G)$ 不包含至少有 $32b/k2^k t^2 n$ 条边的匹配。

证明。一个简单的观察是，我们可以假设上述数值（例如 $32b/k2^k t^2 n$）为整数，因为对于非整数的情况，只需增大 b，直到$32b/k2^k t^2 n$达到下一个整数，即可推出结论。

我们先证明 (i)。在证明 (i) 的过程中，记 j为 $k/2$。我们首先将事件 $H_{k,0}(G)$ 包含一个大星形图，描述为若干特定事件的并集。该事件取决于

- 一个顶点 $u \in V$，
- 是 $W \subseteq V \setminus \{u\}$ 的一个子集，且 $|W| = 32b/k2^k t^2 n$，并且
- 一个子集 $\Gamma \subseteq V \setminus \{u\}$，满足 $|\Gamma| \leqslant 2^j t n$。

对于上述每种选择，我们令 $F(u,W,\Gamma)$ 表示这样一个事件：Γ 是 $G \sim G(n,p)$ 中 u 的邻域，并且

$$e(W,\Gamma) \geqslant 2^k t^2 n |W| .$$

注意，只有当某个三元组 u,W,Γ 满足 $F(u,W,\Gamma)$ 时，(i) 中描述的事件才可能不成立。只需令 u 为星形图的中心，Γ 为其邻域，并令 W 为 $H_{k,0}(G)$ 中 u 的邻点。

对于固定选取的 W 和 Γ，我们有 $\mathbb{E}[e(W,\Gamma)] \leqslant 2^j t^2 n|W| =: \nu$。事件 $F(u,W,\Gamma)$ 意味着 $e(W,\Gamma) \geqslant 2^{k-j}\nu$。因此，根据 (2.6)，我们有：

$$\begin{aligned}
\mathbb{P}\left(F(u,W,\Gamma) | N(u) = \Gamma\right) &\leqslant 2\exp\left(-(k-j)2^{k-j-3}\nu\right) \\
&= 2\exp(-(k-j)2^{k-3}t^2 n|W|) \\
&\leqslant 2\exp(-2b) .
\end{aligned}$$

对于最后一行，我们用了 $|W|$ 的定义以及 $j = k/2$ 这一事实。现在我们将用并集界完成证明。注意，顶点 u 有 n 种选择，且至多有

$$n^{|W|} = \exp\left(\frac{32b\log n}{k2^k t^2 n}\right) \leqslant \exp\left(\frac{32b}{k2^k}\right) \leqslant \exp(b/2)$$

种 W 的选择（我们用了 $k \geqslant 10$）。因此

$$\begin{aligned}\mathbb{P}\left(\bigcup_{u,W,\Gamma} F(u,W,\Gamma)\right) &\leqslant \sum_{u,W,\Gamma} \mathbb{P}\left(N(u)=\Gamma\right)\mathbb{P}\left(F(u,W,\Gamma)|N(u)=\Gamma\right)\\ &\leqslant 2\exp(-2b)\sum_{u,W}\sum_{\Gamma}\mathbb{P}\left(N(u)=\Gamma\right)\\ &\leqslant 2n\exp(-3b/2)\\ &\leqslant \exp(-b)\,,\end{aligned}$$

按要求（我们使用了 $b \geqslant 3\log\, n$）。

(ii) 的界的处理方式完全相同。唯一的区别在于，现在 j 取自区间 $(k/2, \min\{k-6, \log_2(1/t)\}]$，并且 $(k-j)$ 项与 $|W|$ 中的等价项相消。

对于 (iii)，论证类似，不过现在需要考虑一组不相交的点对（一个匹配） $u_1w_1, \dots u_fw_f$，其中 $f := 32b/k2^kt^2n$，因此过程稍微复杂一些。记 W 为 $\{w_1, \dots, w_f\}$。我们还需要考虑集合序列 $\Gamma_1, \dots, \Gamma_f$，而不是单个集合 Γ。令 $F(u_1, \dots, u_f, w_1, \dots, w_f, \Gamma_1, \dots, \Gamma_f)$ 表示以下事件：对于所有 $g = 1, \dots f$，$N(u_g) = \Gamma_g$，并且

$$\sum_{g=1}^{f} |N(w_g) \cap \Gamma_g| \geqslant 2^k t^2 n f\,.$$

显然，事件 $h_{k,0}(G) \geqslant f$ 包含于以下事件的并集中：$F(u_1, \dots, u_f, w_1, \dots, w_f, \Gamma_1, \dots, \Gamma_f)$，其中该并集是对匹配 $u_1w_1, \dots, u_fw_f$以及集合 $\Gamma_1, \dots, \Gamma_f$取并集，且每个集合的基数均不超过 $2^{k/2}tn$。

我们将 $F(u_1, \dots, u_f, w_1, \dots, w_f, \Gamma_1, \dots, \Gamma_f)$ 简记为 $F(\mathbf{u}, \mathbf{w}\ \mathbf{\Gamma})$，并用 $N(\mathbf{u}) = \mathbf{\Gamma}$ 表示满足以下条件的事件：对于所有 $g = 1, \dots, f$，$N(u_g) = \Gamma_g$。

对于每一种这样的选择，$\sum_{g=1}^{f} |N(w_g) \cap \Gamma_g|$ 是若干指示函数之和，用来表示某些边是否包含在 G中。由于边可能出现两次，它不一定服从二项分布，但可以写成两个二项分布之和。因此，正如上面的“星形图”情形一样，由（2.6）可得

$$\mathbb{P}\left(F(\mathbf{u},\mathbf{w},\mathbf{\Gamma})|N(\mathbf{u})=\mathbf{\Gamma}\right) \leqslant 2\exp(-2b)\,.$$

并集界的论证也与 (i) 中给出的类似。至多有

$$n^{2f} \leqslant \exp\left(\frac{32b\log n}{k2^kt^2n}\right) \leqslant \exp(b/2)$$

种匹配的选择（我们使用了 $t^2 n \geqslant \log n$ and $k \geqslant 10$）。因此

$$\begin{aligned}\mathbb{P}\left(\bigcup_{\mathbf{u},\mathbf{w},\mathbf{\Gamma}} F(\mathbf{u},\mathbf{w},\mathbf{\Gamma})\right) &\leqslant \sum_{\mathbf{u},\mathbf{w},\mathbf{\Gamma}} \mathbb{P}\left(N(\mathbf{u})=\mathbf{\Gamma}\right)\mathbb{P}\left(F(\mathbf{u},\mathbf{w},\mathbf{\Gamma})|N(\mathbf{u})=\mathbf{\Gamma}\right)\\ &\leqslant 2\exp(-2b)\sum_{\mathbf{u},\mathbf{w}}\sum_{\mathbf{\Gamma}}\mathbb{P}\left(N(\mathbf{u})=\mathbf{\Gamma}\right)\\ &\leqslant 2\exp(-3b/2)\\ &\leqslant \exp(-b)\,,\end{aligned}$$

如所需。 □

现在很容易推出命题 5.1 中与$h_k(G_i)$相关的部分。

我们首先给出$h_{k,0}(G_i)$的界。将引理 5.3 中“b”的取值设为 $2b$，并取 $p = s = i/N$，可知存在常数 C_1，使得在 $G \sim G(n,s)$中，$H_{k,0}(G)$含有一个由 $C_1 b/k2^k t^2 n$ 条边组成的星形图或匹配的概率至多为 $\exp(-2b)$。$Bin(N,s) = i$ 的概率至少为 n^{-2}，因此该事件在 G_i中的概率至多为 $n^2\exp(-2b)$。所以，除概率至多为 $n^2\exp(-2b)$的情况外，图$H_{k,0}(G_i)$不含大小为$C_1 b/k2^k t^2 n$的星形图或匹配。在这种情况下，$H_{k,0}(G_i)$中的度至多为 $C_1 b/k2^k t^2 n$，且至多有 $2C_1 b/k2^k t^2 n$ 个顶点的度为正，因此根据握手引理，

$$h_{k,0}(G_i) \leqslant \frac{C_1^2 b^2}{k^2 2^{2k} t^4 n^2}\,.$$

现在，我们对 $h_{k,j}(G_i)$ 进行界定，其中 $k/2 < j \leqslant \min\{k-6, \log_2(1/t)\}$。按照上述论证，并使用引理 5.3 的 (ii) 部分，除了概率至多为 $n^2\exp(-2b)$ 的情形外，我们有 $H_{k,j}(G_i)$ 不含大小为 $2C_1 b/(k-j)2^k t^2 n$ 的星形图。此外，$H_{k,j}(G_i)$ 的每条边都与 V_j 相接；根据引理 3.1，除了概率至多为 $\exp(-2b)$ 的情形外，我们有 $|V_j| \leqslant 2^7 b/tnj2^j$。（这假设 $b \geqslant 2tn$；但如果 $b < 2tn$，则可以改用引理 3.1 的第一部分得到 $\Delta(G_i) \leqslant 4tn$，因此特别地，对所有 $j \geqslant 5$，都有 $V_j = \emptyset$。）由此可知，除了概率至多为 $(n^2+1)\exp(-2b)$ 的情形外，我们有

$$h_{k,j}(G_i) \leqslant |V_j|\,\Delta(H_{k,j}(G_i)) \leqslant \left(\frac{2^7 b}{tnj2^j}\right)\left(\frac{2C_1 b}{(k-j)2^k t^2 n}\right) = \frac{2^8 C_1 b^2}{j(k-j)2^{k+j} t^3 n^2}\,.$$

现在，利用这些界和 (5.1)，除了概率至多为 $n^2 \log_2(1/t)\exp(-2b) \leq \exp(-b)$ 的情形外，我们有

$$\begin{aligned} h_k(G) &= \sum_{j\in J} h_{k,j}(G) \\ &\leqslant \frac{2C_1^2 b^2}{k^2 2^{2k} t^4 n^2} + \sum_{j=\lceil k/2\rceil}^{\min\{k-6,\log_2(1/t)\}} \frac{2^8 C_1 b^2}{j(k-j)2^{k+j} t^3 n^2}. \end{aligned}$$

最后的求和至少以公比为 $3/4$ 的几何级数的速度衰减，因此至多是其首项的 4 倍。因此，除概率至多为 $\exp(-b)$ 的情况外，我们有，

$$h_k(G) \leqslant \frac{2C_1^2 b^2}{k^2 2^{2k} t^4 n^2} + \frac{2^{10} C_1 b^2}{\lfloor k/2\rfloor \lceil k/2\rceil 2^{3k/2} t^3 n^2}.$$

随着 k 的范围增至 $2\log_2(1/t)$，我们有 $t^{-1} \geqslant 2^{k/2}$，因此得到界

$$h_k(G) \leqslant \frac{2^{13} C_1^2 b^2}{k^2 2^{2k} t^4 n^2},$$

如所需。

5.2. 控制 $f_k(G_i)$。现在我们将证明命题 5.1 中与$f_k(G_i)$相关的部分。

我们在前一小节中用于控制 $h_{k,\,0}(G_i)$ 的方法也可用于此处。该方法通过限制相关图中星形子图和匹配的大小（在前一种情况下为 $H_{k,\,0}(G)$，而此处为 $F_k(G)$），将对几乎所有 $k \in K_1$ 的取值给出所需结果。

然而，当 k 较小时（当 $2^k \leqslant 2^8\sqrt{\log n}$ 时），这种方法行不通。问题在于，考虑较大的匹配时，并集界论证会失效，因为有 $\exp(\Theta(f\log n))$ 种方法可以选取一个有 f 条边的匹配；而且，正如下文将要看到的，对固定匹配的概率界为 $\exp(-\Theta(2^{2k}f))$。因此，我们对 $k \in K_1^- := \{k \in K_1 : 2^k \leqslant 2^8\sqrt{\log n}\}$ 给出一种略有不同的论证。

我们首先给出对于 k 使得 $2^k \geqslant 2^8\sqrt{\log n}$ 的论证。如同第 5.1 节，我们将界定星形子图和匹配的大小。我们再次在 $G(n,p)$ 中陈述结果。

引理 5.4. 存在一个绝对常数 C，使得以下结论成立。假设 $t \geqslant Cn^{-1/2}(\log n)^{1/2}$，$p \in (0,t)$，且 $b \geqslant 4tn$。设 $k \in K_1$ 满足 $2^k \geqslant 2^8\sqrt{\log n}$，并令 $G \sim G(n,p)$。其概率至少为 $1-\exp(-b)$:

(i) $F_k(G)$ 不包含度为 $Cb/2^{2k}$ 的星形图，(ii) $F_k(G)$ 不包含含有 $Cb/2^{2k}$ 条边的匹配。

证明。我们首先证明（i）。我们将 C 纳入论证，并在稍后确定其值。

我们首先将 $F_k(G)$ 包含一个大型星形图这一事件描述为若干特定事件的并。该事件取决于

- 一个顶点 $u \in V$,
- 一个子集 $W \subseteq V \setminus \{u\}$，其中有 $|W| = Cb/2^{2k+1}$，并且
- 一个子集 $\Gamma \subseteq V \setminus \{u\}$，其中有 $\big||\Gamma| - pn\big| \leqslant 2^{k-4}n^{1/2}$。

对于上述每一种选择，我们定义 $E(u,W,\Gamma)$ 为如下事件：Γ 是 $G \sim G(n,p)$ 中 u 的邻域，并且

$$\big|e(W,\Gamma) - p^2n|W|\big| \geqslant 2^k tn^{1/2}|W|\,.$$

事实上，若 (i) 不成立，则对某个三元组 u,W,Γ，事件 $E(u,W,\Gamma)$ 会发生。只需令 u 为星形图的中心，Γ 为其在 G 中的邻域，并令 W 为 u 在 $F_k(G)$ 中的 $Cb/2^{2k+1}$ 个邻点，且选择这些邻点，使得对于所有 $w \in W$，$D_{uw}(G)$ 的符号都相同。

对于 W 和 Γ 的固定选择，令 $\mu = \mathbb{E}\,[e(W,\Gamma)] = p|W||\Gamma|$，并注意到，根据关于 $|\Gamma|$ 的条件，我们有

$$\big|\mu - p^2n|W|\big| \leqslant 2^{k-4}pn^{1/2}|W| \leqslant 2^{k-4}tn^{1/2}|W|\,.$$

因此，事件 $E(u,W,\Gamma)$ 蕴含

$$\big|e(W,\Gamma) - \mu\big| \geqslant 2^{k-1}tn^{1/2}|W|\,.$$

现在利用 (2.5) 对 $E(u,W,\Gamma)$ 的概率给出界；当然，我们可以使用 $\mu \leqslant 2p^2n|W| \leqslant 2t^2n|W|$，并且对于 $k \in K_1$，我们有 $2^k tn^{1/2} \leqslant 2^{10}t^2n$。因此

$$\begin{aligned}\mathbb{P}\left(E(u,W,\Gamma) \,\middle|\, N(u) = \Gamma\right) &\leqslant \exp\left(\frac{-2^{2k-2}t^2n|W|^2}{4t^2n|W| + 2^k tn^{1/2}|W|}\right)\\ &\leqslant \exp\big(-2^{2k-13}|W|\big)\,.\end{aligned}$$

并集界的论证与第 5.1 节中的类似。选择一对 u,W 至多有 $n^{|W|+1} = \exp((|W|+1)\log n) \leqslant \exp(2|W|\log n)$ 种方式。因此

$$\begin{aligned}\mathbb{P}\left(\bigcup_{u,W,\Gamma} E(u,W,\Gamma)\right) &\leqslant \sum_{u,W,\Gamma} \mathbb{P}\left(N(u) = \Gamma\right)\mathbb{P}\left(E(u,W,\Gamma) \,|\, N(u) = \Gamma\right)\\ &\leqslant \exp\big(-2^{2k-13}|W|\big)\sum_{u,W}\sum_{\Gamma}\mathbb{P}\left(N(u) = \Gamma\right)\\ &\leqslant \exp\big((2\log n - 2^{2k-13})|W|\big)\\ &\leqslant \exp\big(-2^{2k-14}|W|\big)\\ &\leqslant \exp\big(-Cb/2^{14}\big)\,.\end{aligned}$$

注意，在倒数第二个不等式中，我们使用了界 $2^k \geqslant 2^8\sqrt{\log n}$。这证明了所需的界，条件是 $C \geqslant 2^{14}$。

匹配论证可通过类似的论证得出。正如我们在引理 5.3 的证明中所看到的那样。□

对于剩余的情形，当 $2^k \leqslant 2^8\sqrt{\log n}$ 时，我们使用一个略有不同的引理。我们现在不再考虑星形图和匹配，而是考虑若干互不相交的星形图的并。

在剩余的情形中，图 $F_k(G_i)$ 可能非常稠密。我们必须证明的 $f_k(G_i)$ 的上界为 $Cb^2/2^{4k}$；若 $C \geqslant 2^{16}$，则该上界至少为 $b^2/\log n \geqslant n^2/\log n$。下面的引理指出，在相当稠密的图中，我们可以找到少量星形图，它们覆盖相对较多的顶点。

引理 5.5. 设 G 是一个有 n 个顶点且至少有 n^2/r^2 条边的图。那么存在 r 个顶点 $v_1, \ldots, v_r$，属于 G，使得

$$\left|\bigcup_{i=1}^{r} N(v_i)\right| \geqslant \frac{n}{r}.$$

注 5.6。尽管这个引理极其初等，我们起初并未立即找到相关参考资料。如果此前尚未有人研究，那么沿着相同思路寻找“尽可能最优”的结果或许会很有趣。总结性评论中将进一步讨论这一点。

证明。显然，若某个顶点的度至少为 n/r，则结论显然成立，因此我们可以假设所有顶点的度都小于 n/r。

考虑由 G 得到的有向图：将每条边替换为两条有向边（方向相反）。显然，只需找到 v_1、……、v_r，使这些顶点在 D 中的出邻域的并集基数至少为 n/r。注意，D 中所有顶点的入度和出度均至多为 n/r，并且 $e(D) = 2e(G) \geqslant 2n^2/r^2$。

我们可以通过贪心法找到 v_1、……、v_r。令 v_1 为出度最大的顶点，并用 d_1 表示其出度，再令 $S_1 := N^+(v_1)$ 表示 v_1 的出邻点集合。现在，我们从有向图中删除所有指向 S_1 的边。在剩余的有向图中，找到一个出度最大的顶点 v_2，其出度为 d_2，令 $S_2 = S_1 \cup N^+(v_2)$，并删除其他所有指向 S_2 的边。继续这一过程。

在第 i 步，我们可以假设当前集合 S_i 的基数至多为 n/r（否则我们已经完成），因此到目前为止移除的边总数至多为 n^2/r^2。由此可知，至少还剩下 n^2/r^2 条边，因此 $d_i \geqslant n/r^2$。

显然，这一过程至多经过 r 步就会终止，因为 $|S_i| = d_1 + \cdots + d_i \geqslant in/r^2$。□

关于剩余范围 $k \in K_1^-$ 内 $f_k(G_i)$ 的所需界，将由上述结论以及下述关于 $F_k(G)$ 中星形子图并集的界推出。

引理 5.7。存在一个绝对常数 C，使得以下结论成立。设 $t \geqslant Cn^{-1/2}(\log n)^{1/2}$，$p \in (0, t)$，且 $b \geqslant n$。令 $k \in K_1$ 为

使得 $2^k \leqslant 2^8\sqrt{\log n}$，并令 $G \sim G(n,p)$。以至少 $1-\exp(-b)$ 的概率，$F_k(G)$ 中任意由 $r := 2^{2k}$ 个星形子图组成的并集所包含的顶点数都少于 $Cb/2^{2k}$。

我们来看看界 $f_k(G_i) \leqslant Cb^2/2^{4k}$ 是如何由这两个引理推出的。像之前一样（使用 $n^2\exp(-2b) \leqslant \exp(-b)$），将引理 5.7 的结果转用于 $G_i \sim G(n,i)$，则除概率至多为 $\exp(-b)$ 的情形外，我们有：$F_k(G_i)$ 中任意由 $r := 2^{2k}$ 个星形子图构成的并集都包含少于 $C_1b/2^{2k}$ 个顶点，其中 $C_1 \geqslant 1$ 为某个常数。我们断言，这意味着可以得到形如 $f_k(G_i) \leqslant Cb^2/2^{4k}$ 的界。事实上，反设

$$f_k(G_i) \;>\; \frac{C_1^2b^2}{2^{4k}} \;=\; \frac{C_1^2n^2(b/n)^2}{2^{4k}}.$$

那么由引理 5.5 可知，$F_k(G_i)$ 包含 $r' := n2^{2k}/C_1b$ 个顶点，其邻域覆盖至少 $C_1b/2^{2k}$ 个顶点。这与 $r' \leqslant 2^{2k}$ 矛盾。

现在我们已经说明所需的界如何由这些引理得出，接下来以引理 5.7 的证明作为本节的收尾。

该证明与引理 5.3 和 5.4 的证明类似。所考虑的事件可以看作是星形图情形与匹配情形的某种结合。使用以下记号会很有帮助。给定两个集合序列 $\mathbf{A} = (A_1, \ldots, A_r)$ 和 $\mathbf{B} = (B_1, \ldots, B_r)$，我们定义

$$|\mathbf{A}| := \sum_{j=1}^{r} |A_j| \qquad \text{and} \qquad \mathbf{A}\cdot\mathbf{B} := \sum_{j=1}^{r} |A_j||B_j|\,.$$

我们回顾一下第 2 节中的记号：对于两个顶点集合 U, W，我们用 $e(U,W)$ 表示具有重数的边数。我们将此定义推广到顶点子集序列。定义 $\mathbf{A}\cap\mathbf{B} := (A_1\cap B_1, \ldots, A_r\cap B_r)$ 和 $e(\mathbf{A},\mathbf{B}) := \sum_{j=1}^{r} e(A_j,B_j)$。例如，在随机图 $G \sim G(n,p)$ 中，我们有

$$\mathbb{E}\,[e(\mathbf{A},\mathbf{B})] \;=\; p\mathbf{A}\cdot\mathbf{B} \;-\; p|\mathbf{A}\cap\mathbf{B}| \;=\; p\mathbf{A}\cdot\mathbf{B} \;+\; O(p|\mathbf{A}|)\,.$$

引理 5.7 的证明。我们将 C 纳入论证，并稍后固定其取值。我们继续使用 $r := 2^{2k}$。和之前一样，我们把所讨论的事件表示为一个并集。这次我们对以下对象取并集。

- 一个顶点序列 $\mathbf{u} = (u_1, \ldots, u_r)$，以及若干序列的
- 顶点子集 $\mathbf{W} = (W_1, \ldots, W_r)$，使得 $|\mathbf{W}| = Cb/2^{2k+1}$，并且
- 顶点子集 $\boldsymbol{\Gamma} = (\Gamma_1, \ldots, \Gamma_r)$，其中 $\big||\Gamma_j| - pn\big| \leqslant 2^{k-4}n^{1/2}$ 对所有 $j = 1, \ldots, r$。

对于上述每种选择，我们将 $E(\mathbf{u}, \mathbf{W}, \boldsymbol{\Gamma})$ 定义为这样一个事件：对于所有 $j = 1, \ldots, r$，Γ_j 是 $G \sim G(n,p)$ 中 u_j 的邻域，并且即

$$\big|e(\mathbf{W},\boldsymbol{\Gamma}) - p^2n|\mathbf{W}|\big| \;\geqslant\; 2^k t n^{1/2}|\mathbf{W}|\,.$$

如果引理中的事件不成立，即，如果在 $F_k(G)$ 中存在 $r = 2^{2k}$ 个星形子图，它们覆盖了至少 $Cb/2^{2k}$ 个顶点，那么对于某个三元组 $\mathbf{u}, \mathbf{W}, \mathbf{\Gamma}$，事件 $E(\mathbf{u}, \mathbf{W}, \mathbf{\Gamma})$ 就会发生。此外，$\mathbf{W}$ 中的集合可以取为两两不相交，且不与 $\mathbf{u}$ 相交。下面我们来论证这一断言。

首先注意，在这种情况下，这些点对中至少有 $Cb/2^{2k+1}$ 个具有相同符号的偏差。现在，只需令 $\mathbf{u}$ 为这些星形子图的中心，$\mathbf{\Gamma}$ 为它们在 G 中的邻域，并通过令集合 W_j 两两不相交，且从 $F_k(G)$ 中 u_j 的邻点里选取具有所需符号的点，来定义 $\mathbf{W} = (W_1, \ldots, W_r)$。

现在我们来界定事件 $E(\mathbf{u}, \mathbf{W}, \mathbf{\Gamma})$ 的概率。固定 $\mathbf{W}$ 和 $\mathbf{\Gamma}$ 的一种选择，令 $\mu = \mathbb{E}\left[e(\mathbf{W}, \mathbf{\Gamma})\right] = p\mathbf{W} \cdot \mathbf{\Gamma} + O(p|\mathbf{W}|)$。根据三角不等式以及关于 $|\Gamma_j|$ 的界，我们有

$$\left|\mu - p^2 n|\mathbf{W}|\right| \leqslant 2^{k-3} p n^{1/2} |\mathbf{W}| \leqslant 2^{k-3} t n^{1/2} |\mathbf{W}| .$$

因此，事件 $E(\mathbf{u}, \mathbf{W}, \mathbf{\Gamma})$ 意味着

$$\left|e(\mathbf{W}, \mathbf{\Gamma}) - \mu\right| \geqslant 2^{k-1} t n^{1/2} |\mathbf{W}| .$$

现在我们使用（2.5）来界定 $E(\mathbf{u}, \mathbf{W}, \mathbf{\Gamma})$ 的概率；当然，我们可以利用 $\mu \leqslant 2p^2 n|\mathbf{W}| \leqslant 2t^2 n|\mathbf{W}|$，并且对于 $k \in K_1$，有 $2^k t n^{1/2} \leqslant 2^{10} t^2 n$。因此

$$\begin{aligned}
\mathbb{P}\left(E(\mathbf{u}, \mathbf{W}, \mathbf{\Gamma}) | N(\mathbf{u}) = \mathbf{\Gamma}\right) &\leqslant \exp\left(\frac{-2^{2k-2} t^2 n |\mathbf{W}|^2}{4t^2 n|\mathbf{W}| + 2^k t n^{1/2} |\mathbf{W}|}\right) \\
&\leqslant \exp\left(-2^{2k-13}|\mathbf{W}|\right) .
\end{aligned}$$

并集界的论证与之前所使用的类似，不过这次我们需要更加谨慎。选择 $\mathbf{u}$ 的方式至多有 $n^r \leqslant \exp((\log n)^2)$ 种，选择 $\mathbf{W}$ 中元素的方式至多有 $\binom{n}{|\mathbf{W}|} \leqslant \exp\left(|\mathbf{W}| \log(en/|\mathbf{W}|)\right)$ 种，而将这些元素分配给集合 W_1、……、W_r 的方式有 $r^{|\mathbf{W}|}$ 种。因此，二元组 $\mathbf{u}$、$\mathbf{W}$ 的选择总数至多为

$$\exp\left((\log n)^2 + |\mathbf{W}| \log\left(\frac{enr}{|\mathbf{W}|}\right)\right) \leqslant \exp\left((\log n)^2 + 4k|\mathbf{W}|\right) .$$

因此

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{\mathbf{u},\mathbf{W},\mathbf{\Gamma}} E(\mathbf{u}, \mathbf{W}, \mathbf{\Gamma})\right) &\leqslant \sum_{\mathbf{u},\mathbf{W},\mathbf{\Gamma}} \mathbb{P}\left(N(\mathbf{u}) = \mathbf{\Gamma}\right) \mathbb{P}\left(E(\mathbf{u}, \mathbf{W}, \mathbf{\Gamma}) | N(\mathbf{u}) = \mathbf{\Gamma}\right) \\
&\leqslant \exp\left(-2^{2k-13}|\mathbf{W}|\right) \sum_{u,W} \sum_{\mathbf{\Gamma}} \mathbb{P}\left(N(\mathbf{u}) = \mathbf{\Gamma}\right) \\
&\leqslant \exp\left((\log n)^2 + 4k|\mathbf{W}| - 2^{2k-13}|\mathbf{W}|\right) \\
&\leqslant \exp\left(-2^{2k-14}|\mathbf{W}|\right) \\
&\leqslant \exp\left(-Cb/2^{14}\right) .
\end{aligned}$$

只要满足 $C \geqslant 2^{14}$，这就证明了所需的界。 □

5.3. 增量 $X_{\triangle}(G_i)$。现在我们推导出关于增量的以下界 s.

引理 5.8. 存在一个绝对常数 C，使得以下结论成立。设$Cn^{-1/2}(\log n)$ $^{1/2} \leqslant t \leqslant 1/2$。对于所有 $n \leqslant b \leqslant t^2n^2$ 和 $i \leqslant m$，我们有

$$\mathbb{P}\left(\mathbb{E}\left[X_{\triangle}(G_i)^2|G_{i-1}\right] \geqslant Cbt^2\right) \leqslant \exp(-b)$$

并且，对于 $b > t^2n^2$，我们有

$$\mathbb{P}\left(\mathbb{E}\left[X_{\triangle}(G_i)^2|G_{i-1}\right] \geqslant \frac{Cb^2}{n^2}\right) \leqslant \exp(-b)\,.$$

证明。该证明依赖于将 $X_{\triangle}(G_i)$ 与共度偏差联系起来。根据 [20] 中引理 5.4 的证明，我们得到

$$X_{\triangle}(G_i) \;=\; 6D_{uw}(G_{i-1}) \;-\; \mathbb{E}\left[6D_{e_i}(G_{i-1}) \;\middle|\; G_{i-1}\right] \tag{5.2}$$

其中，$e_i = uw$ 是选作第 i 条边的顶点对。由于在给定 G_{i-1} 的条件下，$X_{\triangle}(G_i)$ 的条件均值为 0，因此我们有

$$\begin{aligned}\mathbb{E}\left[X_{\triangle}(G_i)^2 \mid G_{i-1}\right] &= \mathrm{Var}(X_{\triangle}(G_i) \mid G_{i-1}) \\ &= 36\mathrm{Var}(D_{uw}(G_{i-1}) \mid G_{i-1}) \leqslant 36\mathbb{E}\left[D_{uw}(G_{i-1})^2 \mid G_{i-1}\right]\,.\end{aligned}$$

因此，

$$\begin{aligned}\mathbb{E}\left[X_{\triangle}(G_i)^2|G_{i-1}\right] &\leqslant 36\frac{1}{N-i+1}\sum_{uw\notin G_{i-1}} D_{uw}(G_{i-1})^2 \\ &\leqslant \frac{144}{n^2}\sum_{u} D_u(G_{i-1})^2\,.\end{aligned}$$

这两个不等式现在可直接由推论 5.2 得出。 □

6. Triangle counts in $G(n,m)$ – upper bounds on deviations

本节中，我们证明 $G(n,m)$ 中三角形计数偏差概率的上界。也就是说，我们证明定理 1.1 的上界部分。下面我们陈述一个蕴含定理 1.1 这一部分的结果（命题 6.1）。

为了陈述命题 6.1，我们先回顾引言中定义的函数 NORMAL(b,t)、STAR(b,t)、HUB(b,t)、CLIQUE(b,t) 和 $M(b,t)$。我们回顾

that

$$\begin{aligned}
\text{NORMAL}(b,t) &:= b^{1/2}t^{3/2}n^{3/2}\,,\\
\text{STAR}(b,t) &:= \frac{b^2t}{\ell^2}\,1_{b\leqslant n\ell}\,,\\
\text{HUB}(b,t) &:= \frac{btn}{\ell}\,1_{b\geqslant n\ell}\,,\\
\text{CLIQUE}(b,t) &:= \frac{b^{3/2}}{\ell^{3/2}} \quad \text{and}\\
M(b,t) &:= \max\{\text{NORMAL}(b,t), \text{STAR}(b,t), \text{HUB}(b,t), \text{CLIQUE}(b,t)\}\,.
\end{aligned}$$

我们将给出 $M(b,t)$ 的一些下界，这些结果在本节后续部分会派上用场。首先，我们断言，

$$M(b,t) \geqslant bt^{3/2}n\,. \tag{6.1}$$

对所有 $b \geqslant 1$ 均成立。这很容易验证，因为对于 $b \leqslant n$，我们有 NORMAL(b,t) $\geqslant bt^{3/2}n$；对于 $n < b \leqslant n\ell$，我们有星形图$(b,t) \geqslant bt^{3/2}n$；对于 $n\ell < b \leqslant tn^2\ell$，我们有枢纽$(b,t) \geqslant bt^{3/2}n$。特别地，(6.1) 意味着

$$M(b,t) \geqslant bt^2n\,. \tag{6.2}$$

我们还断言，

$$M(b,t) \geqslant t\kappa(b,t) \tag{6.3}$$

对于所有 $b \geqslant 32tn$，其中 $\kappa(b,t)$ 的定义见第3节。为验证这一点，注意到只要 $32tn \leqslant b < t^{1/2}n\ell$，NORMAL$(b,t) = b^{1/2}t^{3/2}n^{3/2} \geqslant t^2n^2 = t\kappa(b,t)$。此外，只要 $t^{1/2}n\ell \leqslant b < n\ell$，星形图$(b,t) = t\kappa(b,t)$，且若$n\ell < b \leqslant tn^2\ell$，枢纽$(b,t) = t\kappa(b,t)$。

我们回顾第 2.3 节中的三角形计数偏差 $D_\triangle(G_m)$ 的鞅表示：

$$D_\triangle(G_m) = \sum_{i=1}^{m}\left[3\,\frac{(N-m)_2(m-i)}{(N-i)_3}\,X_\wedge(G_i) + \frac{(N-m)_3}{(N-i)_3}\,X_\triangle(G_i)\right].$$

现在我们可以陈述本节将证明的上界结果。

命题 6.1. 存在一个绝对常数 C 使得以下结论成立。对于所有 $Cn^{-1/2}(\log n)^{1/2} \leqslant t \leqslant 1/2$ 以及 $3\log n \leqslant b \leqslant tn^2\ell$，我们有

$$\mathbb{P}\left(D_\triangle(G_m) \geqslant CM(b,t)\right) \leqslant \exp(-b)\,.$$

我们使用分治鞅方法来证明命题 6.1。好奇的读者可能会想，为什么我们不能直接使用直接鞅方法，通过 Freedman 不等式来界定出现这种偏差的概率。问题在于，在这一过程中偶尔可能会出现一些

非常大的增量。如果让这些增量出现在鞅论证中，我们得到的界就会更弱。因此，我们单独考虑较大的增量。

现在我们引入增量$X_\wedge(G_i)$和 $X_\triangle(G_i)$的截断版本$X'_\wedge(G_i)$和$X'_\triangle(G_i)$。截断这些增量所取的值将是

$$K_\wedge \;=\; K_\wedge(b,t) \;:=\; 2^8\,\frac{M(b,t)}{bt}$$

and

$$K_\triangle \;=\; K_\triangle(b,t) \;:=\; 2^{16}\,\frac{M(b,t)}{b}$$

分别。我们设定

$$X'_\wedge(G_i) \;=\; X_\wedge(G_i)1_{X_\wedge(G_i)\leqslant K_\wedge} \qquad \text{and} \qquad X'_\triangle(G_i) \;=\; X_\triangle(G_i)1_{X_\triangle(G_i)\leqslant K_\triangle}\,.$$

我们还将考虑随机变量

$$Z_\wedge(G_i) \;=\; X_\wedge(G_i)1_{X_\wedge(G_i)> K_\wedge} \qquad \text{and} \qquad Z_\triangle(G_i) \;=\; X_\triangle(G_i)1_{X_\triangle(G_i)> K_\triangle}\,.$$

注意，$X_\wedge(G_i) = X'_\wedge(G_i) + Z_\wedge(G_i)$，且 $X_\triangle(G_i) = X'_\triangle(G_i) + Z_\triangle(G_i)$。

现在我们可以将偏差 $D_\triangle(G_m)$ 表示为 $D_\triangle(G_m) = D'_\triangle(G_m) + N^*_\triangle(G_m)$，其中

$$D'_\triangle(G_m) \;=\; \sum_{i=1}^{m}\left[3\,\frac{(N-m)_2(m-i)}{(N-i)_3}\,X'_\wedge(G_i) \;+\; \frac{(N-m)_3}{(N-i)_3}\,X'_\triangle(G_i)\right] \tag{6.4}$$

and

$$N^*_\triangle(G_m) \;=\; \sum_{i=1}^{m}\left[3\,\frac{(N-m)_2(m-i)}{(N-i)_3}\,Z_\wedge(G_i) \;+\; \frac{(N-m)_3}{(N-i)_3}\,Z_\triangle(G_i)\right].$$

由于 Z 随机变量均为非负数，且系数分别至多为 $3t$ 和 1，因此我们有

$$N^*_\triangle(G_m) \;\leqslant\; \sum_{i=1}^{m}\left[3t\,Z_\wedge(G_i) \;+\; Z_\triangle(G_i)\right]. \tag{6.5}$$

我们将 $D_\triangle(G_m)$ 表示为 $D'_\triangle(G_m)$ 与 $N^*_\triangle(G_m)$ 之和。其中，$D'_\triangle(G_m)$ 是一个超鞅（因为截断只会使其值减小），而 $N^*_\triangle(G_m)$ 计数了“大”增量的贡献。为了完成命题 6.1 的证明，显然只需证明存在绝对常数 C_1 和 C_2，使得对于所有满足 $\max\{C_1,C_2\}n^{-1/2}(\log n)^{1/2} \leqslant t \leqslant 1/2$ 且 $3\log n \leqslant b \leqslant tn^2\ell$ 的情形，我们都有

$$\mathbb{P}\left(D'_\triangle(G_m) \;\geqslant\; C_1M(b,t)\right) \;\leqslant\; \exp(-2b) \tag{6.6}$$

and

$$\mathbb{P}\left(N^*_\triangle(G_m) \;\geqslant\; C_2M(b,t)\right) \;\leqslant\; \exp(-2b)\,. \tag{6.7}$$

我们将在接下来的两个小节中证明这两个结论。

6.1. 控制 $D'_{\triangle}(G_m)$。在本小节中，我们证明 (6.6)。由于 (6.4) 将$D'_{\triangle}(G_m)$表示为一个增量为

$$\mathbb{X}'_i := 3\,\frac{(N-m)_2(m-i)}{(N-i)_3}\,X'_{\wedge}(G_i) + \frac{(N-m)_3}{(N-i)_3}\,X'_{\triangle}(G_i)\,,$$

我们可以使用 Freedman 不等式来界定 $D'_{\triangle}(G_m) \geqslant C_1 M(b,t)$ 的概率。我们所需要的只是对增量 $\mathbb{X}'_i$ 进行一定控制。由于系数分别至多为$3t$和 1，因此我们有

$$|\mathbb{X}'_i| \leqslant 3t|X'_{\wedge}(G_i)| + |X'_{\triangle}(G_i)| \tag{6.8}$$

确定性地。由于截断限制了这些绝对值，因此我们有 $|\mathbb{X}'_i| \leqslant 2^{17} M(b,t)/b$ 确定性地。

另一方面，利用 (6.8) 以及 $(x+y)^2 \leqslant 2x^2 + 2y^2$ 这一事实，我们有

$$|\mathbb{X}'_i|^2 \leqslant 18t^2\,|X'_{\wedge}(G_i)|^2 + 2|X'_{\triangle}(G_i)|^2\,. \tag{6.9}$$

这个界与下面的引理结合起来，将有助于控制该过程的二次变差。其证明将用到（6.9）以及引理 3.4 和 5.8，它们分别给出了增量 $X_{\wedge}(G_i)$ 和 $X_{\triangle}(G_i)$ 的条件二阶矩的界。

引理 6.2。存在一个绝对常数 C，使得以下结论成立。设$Cn^{-1/2}(\log n)^{1/2} \leqslant t \leqslant 1/2$。对于 $n \leqslant b \leqslant t^2n^2$ 和 $i \leqslant m$，有

$$\mathbb{P}\left(\mathbb{E}\left[(\mathbb{X}'_i)^2|G_{i-1}\right] \geqslant Cbt^2\right) \leqslant \exp(-b)$$

并且，对于 $b > t^2n^2$，有

$$\mathbb{P}\left(\mathbb{E}\left[(\mathbb{X}'_i)^2|G_{i-1}\right] \geqslant \frac{Cb^2}{n^2}\right) \leqslant \exp(-b)\,.$$

此外，对于所有 $b \geqslant n$,

$$\mathbb{P}\left(\sum_{i=1}^{m}\mathbb{E}\left[(\mathbb{X}'_i)^2|G_{i-1}\right] \geqslant C(bt^3n^2 + b^2t)\right) \leqslant \exp(-b)\,.$$

证明。我们首先指出，引理 3.4 和 5.8 中的界对$X'_{\wedge}(G_i)$ 和 $X'_{\triangle}(G_i)$ 也成立，因为截断只能减小方差。

现在，前两个界由 (6.9)、引理 3.4 和 5.8，以及以下事实立即得出：

$$t^2\kappa(b,t) \leqslant \max\{bt^2, b^2/n^2\}$$

在整个范围 $b \geqslant n$ 内。

进一步的结论可由并集界轻易得到。注意，该界等于另外两个界之和再乘以步数 $m \leqslant tn^2$。□

现在我们可以证明 (6.6) 了。我们将对末值为 $D'_{\triangle}(G_m)$ 的超鞅 (6.4) 应用 Freedman 不等式。我们已经证明，增量 $\mathbb{X}'_i$ 确定性地满足 $|\mathbb{X}'_i| \leqslant 2^{17}M(b,t)/b$。现在根据 b 的取值分为两种情况。

对于 $3\log n \leqslant b \leqslant n$，我们设定 $\beta = 8C't^3n^3$，其中 $C' \geqslant 1$ 是引理 6.2 的常数，使得 $\mathbb{P}\left(\sum_{i=1}^m \mathbb{E}\left[(\mathbb{X}'_i)^2|G_{i-1}\right] \geqslant \beta\right) \leqslant \exp(-2n) \leqslant \exp(-2b)$。另请注意，我们可以选择一个常数 $C_1 \geqslant 16$，使得 $\beta \leqslant C_1\mathrm{NORMAL}(b,t)^2/b \leqslant C_1M(b,t)^2/b$。

由 Freedman 不等式（引理 2.2）可知，

$$\begin{aligned}
\mathbb{P}\left(D'_{\triangle}(G_m) \geqslant C_1M(b,t)\right) &\leqslant \exp\left(\frac{-C_1^2M(b,t)^2}{2\beta + 2C_1M(b,t)(2M(b,t)/b)}\right) + \exp(-2b)\\
&\leqslant \exp\left(\frac{-C_1^2M(b,t)^2}{2C_1M(b,t)^2/b + 2C_1M(b,t)(2M(b,t)/b)}\right) + \exp(-2b)\\
&\leqslant \exp\left(\frac{-C_1b}{6}\right) + \exp(-2b)\\
&\leqslant 2\exp(-2b)\\
&\leqslant \exp(-b)\,.
\end{aligned}$$

在 $b > n$ 的情况下，论证过程类似。现在我们令 $\beta = 4C'(bt^3n^2 + b^2t)$，从而使得 $\mathbb{P}\left(\sum_{i=1}^m \mathbb{E}\left[(\mathbb{X}'_i)^2|G_{i-1}\right] \geqslant \beta\right) \leqslant \exp(-2b)$。还需注意，我们可以选择一个常数 $C_1 \geqslant 16$，使得对于所有 $b \geqslant n$，均有 $\beta \leqslant C_1M(b,t)^2/b$。现在，利用 Freedman 不等式，按照上述完全相同的不等式推导步骤，即可得到所需的界。这就完成了对 (6.6) 的证明。

6.2. 控制 $N^*_{\triangle}(G_m)$。在本小节中，我们通过证明界（6.7）完成命题 6.1 的证明。也就是说，我们证明对于某个常数 C_2，不等式

$$\mathbb{P}\left(N^*_{\triangle}(G_m) \geqslant C_2M(b,t)\right) \leqslant \exp(-2b)$$

对所有 $Cn^{-1/2}(\log n)^{1/2} \leqslant t \leqslant 1/2$ 和 $3\log n \leqslant b \leqslant tn^2\ell$ 均成立。正如我们上面在 (6.5) 中所看到的，我们有

$$N^*_{\triangle}(G_m) \leqslant 3\sum_{i=1}^m t\,Z_{\wedge}(G_i) + \sum_{i=1}^m Z_{\triangle}(G_i)\,.$$

因此，根据 (2.9)，即三角不等式和并集界，只需证明存在绝对常数 C_3 和 C_4 使得

$$\mathbb{P}\left(\sum_{i=1}^m t\,Z_{\wedge}(G_i) \geqslant C_3\,M_{\triangle}(b,t)\right) \leqslant \exp(-3b) \tag{6.10}$$

and

$$\mathbb{P}\left(\sum_{i=1}^{m} Z_{\triangle}(G_i) \geqslant C_4\, M_{\triangle}(b,t)\right) \leqslant 2\exp(-3b)\,. \tag{6.11}$$

我们将在以下小节中证明 (6.10) 和 (6.11)。

6.2.1. 控制 $\sum_{i=1}^{m} t\, Z_{\wedge}(G_i)$。在本小节中，我们证明 (6.10)。证明与第4.2节的论证相似。同样，我们的主要目标是将求和与度偏差平方和联系起来，从而可以利用命题 3.2 推出 (6.10)。首先，对于 $i \leqslant m$，我们有

$$X_{\wedge}(G_i) \leqslant A_{\wedge}(G_i) = 2\sum_{v\in e_i} d_v(G_{i-1}) \leqslant 2\sum_{v\in e_i} d_v(G_m) \leqslant 4\max_{v\in e_i} d_v(G_i)\,.$$

因此，事件 $X_{\wedge}(G_i) > K_{\wedge}$ 只有在两个顶点 $v \in e$ 中有一个满足 $d_v(G_m) > K_{\wedge}/4$ 时才会发生。由此可得

$$Z_{\wedge}(G_i) = X_{\wedge}(G_i)1_{X_{\wedge}(G_i)>K_{\wedge}} \leqslant 4\sum_{v\in e_i} d_v(G_m)1_{d_v(G_m)>K_{\wedge}/4}\,. \tag{6.12}$$

将此表达式对 $i \leqslant m$ 求和，我们得到上界

$$\begin{aligned}\sum_{i=1}^{m} Z_{\wedge}(G_i) &\leqslant 4\sum_{v\in V} d_v(G_m)1_{d_v(G_m)>K_{\wedge}/4} \sum_{e\in E(G_m)} 1_{v\in e} \\ &= 4\sum_{v\in V} d_v(G_m)^2 1_{d_v(G_m)>K_{\wedge}/4}\,.\end{aligned} \tag{6.13}$$

回顾 (6.2)，根据定义，$K_{\wedge} = 2^8 M(b,t)/(bt) \geqslant 2^8 tn$。特别地，$K_{\wedge}/4 \geqslant 64tn$，因此，当 $\Delta(G_m) \leqslant 64tn$ 时，上述求和为 0。

注意，应用引理 3.1，并令“b”为 $96tn$，可得

$$\mathbb{P}\left(\Delta(G_m) > 64tn\right) \leqslant \mathbb{P}\left(\Delta(G_m) > \frac{96tn}{\log(96/e)}\right) \leqslant \exp(-96tn)\,. \tag{6.14}$$

特别地，若 $b \leqslant 32tn$，则 $\sum_{i=1}^{m} Z_{\wedge}(G_i)$ 非零的概率至多为 $\exp(-3b)$，此时 (6.10) 得证。

现在假设 $b \geqslant 32tn$。注意，对所有满足$d_v(G_m) \geqslant 32tn$的顶点 v，都有 $d_v(G_m) \leqslant 2D_v(G_m)$，因此由 (6.13) 可得

$$\sum_{i=1}^{m} tZ_{\wedge}(G_i) \leqslant 16t\sum_{v\in V} D_v(G_m)^2$$

确定地。现在我们利用命题 3.2 来界定该和较大的概率。回顾一下，(6.3) 中，$t\kappa(b,t) \leqslant M(b,t)$。因此，根据命题 3.2，

存在一个常数 C_3，使得

$$\mathbb{P}\left(\sum_{i=1}^{m} tZ_{\wedge}(G_i) \geqslant C_3\, M(b,t)\right) \leqslant \mathbb{P}\left(\sum_{i=1}^{m} tZ_{\wedge}(G_i) \geqslant C_3\, t\kappa(b,t)\right) \leqslant \exp(-3b)\,.$$

这就完成了我们对（6.10）的证明。

6.2.2. 控制 $\sum_{i=1}^{m} Z_{\triangle}(G_i)$。在本节中，我们证明 (6.11)。回顾一下，$Z_{\triangle}(G_i)$ 定义为

$$Z_{\triangle}(G_i) \;=\; X_{\triangle}(G_i)1_{X_{\triangle}(G_i)>K_{\triangle}}$$

其中 $K_{\triangle} = 2^{16}M(b,t)/b$。我们的目标是证明：存在某个常数 C_4，使得我们有

$$\mathbb{P}\left(\sum_{i=1}^{m} Z_{\triangle}(G_i) \;>\; C_4 M_{\triangle}(b,t)\right) \;\leqslant\; 2\exp(-3b)$$

对所有 $Cn^{-1/2}(\log\, n)^{1/2} \leqslant t \leqslant 1/2$ 以及 $3\log n \leqslant b \leqslant tn^2\ell$。

对于每个 $i \leqslant m$，我们可以根据第 i 步选取的边的共度来控制 $X_{\triangle}(G_i)$。具体来说，

$$X_{\triangle}(G_i) \;\leqslant\; A_{\triangle}(G_i) \;=\; 6d_{e_i}(G_{i-1})\,.$$

因此，要使事件 $X_{\triangle}(G_i) > K_{\triangle}$ 发生，必须满足 $d_{e_i}(G_{i-1}) > K_{\triangle}/6$。由此可得

$$\sum_{i=1}^{m} Z_{\triangle}(G_i) \;\leqslant\; 6\sum_{i=1}^{m} d_{e_i}(G_{i-1})1_{\{d_{e_i}(G_{i-1})>K_{\triangle}/6\}}\,. \tag{6.15}$$

我们考虑对总和的两种不同贡献类型，具体取决于共度是否可由度偏差“解释”。对于过程中加入的每条边 $e_i = uv$，若其具有 $d_{e_i}(G_{i-1}) > K_{\triangle}/6$，则如果存在某个时刻 $i-1 \leqslant j \leqslant m$，使得 e_i 为良好，便将其归类为良好。

$$d_{e_i}(G_j) \;\leqslant\; 64t(d_u(G_j) + d_v(G_j))\,.$$

否则 e_i 是坏的。

As

$$\sum_{i=1}^{m} Z_{\triangle}(G_i) \;\leqslant\; 6\sum_{i=1}^{m} d_{e_i}(G_{i-1})1_{e_i \text{ is good}} + 6\sum_{i=1}^{m} d_{e_i}(G_{i-1})1_{e_i \text{ is bad}}\,, \tag{6.16}$$

我们已将目标归结为证明以下两个引理（模去 (2.9)），即通常的并集界和三角形不等式。

引理 6.3. 存在一个绝对常数 C，使得以下结论成立。假设 $Cn^{-1/2}(\log n)^{1/2} \leqslant t \leqslant 1/2$ 且 $3\log n \leqslant b \leqslant tn^2\ell$。那么，除概率至多为 $\exp(-3b)$ 的情况外，

$$\sum_{i=1}^{m} d_{e_i}(G_{i-1})1_{\{e_i \text{ is good}\}} \leqslant CM(b,t)\,.$$

引理 6.4. 存在一个绝对常数 C，使得以下结论成立。假设 $Cn^{-1/2}(\log n)^{1/2} \leqslant t \leqslant 1/2$ 且 $3\log n \leqslant b \leqslant tn^2\ell$。那么，除概率至多为 $\exp(-3b)$ 外，

$$\sum_{i=1}^{m} d_{e_i}(G_{i-1})1_{\{e_i \text{ is bad}\}} \leqslant CM(b,t)\,.$$

我们从引理 6.3 开始。

引理 6.3 的证明。由于在过程中度只会增加，对于每条好边 e_i，我们有 $K_\triangle/6 \leqslant d_{e_i}(G_{i-1}) \leqslant 64t(d_u(G_m) + d_v(G_m))$。

特别地，如果 e_i 是良好的，那么 $d_v(G_m) \geqslant K_\triangle/2^{10}t$ 对两个顶点 $v \in e_i$ 中的一个成立。因此。

$$d_{e_i}(G_{i-1})1_{\{e_i \text{ is good}\}} \leqslant 128t \sum_{v\in e_i} d_v(G_m)1_{d_v(G_m)>K_\triangle/2^{10}t}\,. \tag{6.17}$$

因此

$$\begin{aligned}\sum_{i=1}^{m} d_{e_i}(G_{i-1})1_{\{e_i \text{ is good}\}} &\leqslant 128t \sum_{v\in V} d_v(G_m)1_{d_v(G_m)>K_\triangle/2^{10}t} \sum_{e\in E(G_m)} 1_{v\in e} \\ &= 128t \sum_{v\in V} d_v(G_m)^2 1_{d_v(G_m)>K_\triangle/2^{10}t}.\end{aligned}$$

我们还回顾一下（6.2），即 $K_\triangle/2^{10}t = 64M(b,t)/bt \geqslant 64tn$。特别地，如果 $\Delta(G_m) \leqslant 64tn$，那么上面的最后一个求和自动为 0。正如我们在上文（6.14）中所见，我们有 $\mathbb{P}(\Delta(G_m) > 64tn) \leqslant \exp(-96tn)$。特别地，如果 $b < 32tn$，那么 $\sum_{i=1}^{m} d_{e_i}(G_{i-1})1_{\{e_i \text{ 是良好的}\}}$ 非零的概率至多为 $\exp(-3b)$。

现在我们可以假设 $b \geqslant 32tn$。由于对于所有满足 $d_v(G_m) \geqslant 64tn$ 的 v，都有 $d_v(G_m) \leqslant 2D_v(G_m)$，因此可得

$$128t \sum_{v\in V} d_v(G_m)^2 1_{d_v(G_m)>K_\triangle/2^9 t} \leqslant 256t \sum_{v\in V} D_v(G_m)^2\,.$$

现在我们将使用命题 3.2 来界定该和较大的概率。我们回顾式 (6.3)，其表明 $t\kappa(b,t) \leqslant CM(b,t)$。现在，令 C 为 2^{10} 倍

命题 3.2 的常数，我们有

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^{m} d_{e_i}(G_{i-1})1_{\{e_i \text{ is good}\}} > CM(b,t)\right) &\leqslant \mathbb{P}\left(\sum_{i=1}^{m} d_{e_i}(G_{i-1})1_{\{e_i \text{ is good}\}} > Ct\kappa(b,t)\right)\\ &\leqslant \mathbb{P}\left(256t\sum_{v\in V} D_v(G_m)^2 > Ct\kappa(b,t)\right)\\ &\leqslant \mathbb{P}\left(\sum_{v\in V} D_v(G_m)^2 > \frac{C}{256}\kappa(b,t)\right)\\ &\leqslant \exp(-3b)\,.\end{aligned}$$

□

现在我们来界定坏边的共度之和。为此，我们证明：若 e_i 是坏边，则它属于第 5 节中定义的、由共度较大的点对构成的图 $H_k(G_m)$ 之一。利用第 5 节中证明的界，我们就能得到这类边数量的上界。这对于可能共度值的较大范围已足够。

对于仅略大于 $K_\triangle$、事实上一直到 $K_\triangle \ell$ 的共度值，则需要更仔细的分析。我们会用到前面提及的界，但还需考虑在此过程中选中大量这类边的概率。为此，我们采用一种论证，将该过程与一个具有独立增量的过程耦合，而后者服从二项分布。

引理 6.4 的证明。回顾一下，我们希望找到一个绝对常数 C，使得对于所有 $Cn^{-1/2}(\log n)^{1/2} \leqslant t \leqslant 1/2$ 以及所有 $3\log n \leqslant b \leqslant tn^2\ell$，都有

$$\mathbb{P}\left(\sum_{i=1}^{m} d_{e_i}(G_{i-1})1_{\{e_i \text{ is bad}\}} > CM(b,t)\right) \leqslant \exp(-3b)\,. \tag{6.18}$$

正如我们在证明前的讨论中提到的，我们将分别考虑共度介于 $K_\triangle/6$ 和 $K_\triangle\ell$ 之间，以及大于$K_\triangle\ell$ 的共度所作出的贡献。我们将把证明归结为以下陈述：对于某个常数 $C > 0$，以下两个断言成立：

$$\mathbb{P}\left(\sum_{i=1}^{m} d_{e_i}(G_{i-1})1_{\{e_i \text{ is bad}, d_{e_i}(G_{i-1})>K_\triangle\ell\}} > CM(b,t)\right) \leqslant \exp(-4b) \tag{6.19}$$

and

$$\mathbb{P}\left(\sum_{i=1}^{m} d_{e_i}(G_{i-1})1_{\{e_i \text{ is bad}, d_{e_i}(G_{i-1})\in[K_\triangle/6, K_\triangle\ell)\}} > CM(b,t)\right) \leqslant 2\exp(-4b)\,. \tag{6.20}$$

事实上，若 (6.19) 和 (6.20) 均成立，则显然，除概率至多为 3exp(−4b) 6 exp(−3b) 的情形外，我们有

$$\sum_{i=1}^{m} d_{e_i}(G_{i-1}) 1_{\{e_i \text{ is bad}\}} \leqslant 2CM(b,t)\,.$$

证明 (6.19) 相对直接。我们将利用第5节的结果来界定共度较大的点对数量。这些界足够强，即使选取每一条共度如此大的边，其贡献仍至多为 $CM(b,t)$。

回顾第5节，我们定义了集合

$$H_k(G) := \{uw : d_{uw}(G) \in [2^k t^2 n, 2^{k+1} t^2 n)\,,\, d_u(G), d_w(G) \leqslant 2^{k-5} tn\}$$

对于 $k \in K_2 := \{10, \ldots, \lfloor 2\log_2(1/t) \rfloor\}$。并且我们设 $h_k(G) = |H_k(G)|$。

由（6.2）以及 $K_\triangle$ 定义为 $2^{16}M(b,t)/b$，可得 $K_\triangle/6 \geqslant 2^{10}t^2 n$。令 $k_0 \geqslant 10$ 为满足 $2^k t^2 n \leqslant K_\triangle/6$ 的最大 k，令 k_1 为满足 $2^k t^2 n \leqslant K_\triangle \ell$ 的最大 k。另需注意，由于 $K_\triangle = 2^{16}M(b,t)/b \geqslant 2^{16} t^{3/2} n$，根据（6.1），我们有

$$k_1 \geqslant k_0 \geqslant k_0 - 12 \geqslant \log_2(t^{-1/2}) \geqslant \ell/2\,. \tag{6.21}$$

断言：如果 e_i 是坏的且 $d_{e_i}(G_{i-1}) \geqslant K_\triangle \ell$，那么 $e_i \in \bigcup_{k=k_1}^{2\log_2(1/t)} H_k(G_m)$。

断言的证明：首先，由于共度在整个过程中单调不减

$$d_{e_i}(G_m) \geqslant d_{e_i}(G_{i-1}) \geqslant K_\triangle \ell \geqslant 2^{k_1} t^2 n\,.$$

因此，$d_{e_i}(G_m)$ 属于某个集合 $H_k(G)$ 的范围，其中 $k \geqslant k_1$。

剩下要证明的只有，$H_k(G)$ 定义中的第二个条件成立。为此，只需有 $d_{e_i}(G_m) > 64t(d_u(G_m) + d_v(G_m))$，这由 “ 坏 ” 的定义立即可得。断言得证。

根据命题 5.1，除概率至多为 exp(−4b) 的情况外，对于某个常数 C_1，我们有

$$h_k(G_i) \leqslant \frac{C_1 b^2}{k^2 2^{2k} t^4 n^2} \qquad \text{for all } k \in K_2 \text{ and } i \leqslant m\,.$$

假设这些界成立，并证明（6.19）中和式所需的界。由于对于边 $e_i \in H_k(G_m)$ 有 $d_{e_i}(G_{i-1}) \leqslant d_{e_i}(G_m) \leqslant 2^{k+1} t^2 n$，除非有

概率至多为 $\exp(-4b)$,

$$\begin{aligned}\sum_{i=1}^{m} d_{e_i}(G_{i-1})1_{\{e_i \text{ is bad}, d_{e_i}(G_{i-1})>K_\triangle \ell\}} &\leqslant \sum_{k=k_1}^{2\log_2(1/t)} 2^{k+1}t^2 n \frac{C_1 b^2}{k^2 2^{2k} t^4 n^2} \\ &\leqslant \sum_{k=k_1}^{2\log_2(1/t)} \frac{2C_1 b^2}{k^2 2^k t^2 n} \\ &\leqslant \frac{4C_1 b^2}{k_1^2 2^{k_1} t^2 n} \\ &\leqslant \frac{8C_1 b^2}{k_1^2 K_\triangle \ell} \\ &\leqslant \frac{2^{21} C_1 b^3}{\ell^3 M(b,t)} \qquad (6.22) \\ &\leqslant 2^{21} C_1 M(b,t)\,.\end{aligned}$$

注意，倒数第二个不等式使用了 $K_\triangle$ 的定义以及上文观察到的不等式 $k_1 \geqslant \ell/2$。最后一个不等式则利用了 $M(b,t)^2 \geqslant$ 团$(b,t)^2 = b^3/\ell^3$。至此，(6.19) 的证明完成。

现在考虑 (6.20)。这一次需要采用更微妙的论证。我们仍将使用基数 $h_k(G_{i-1})$ 的界，但还必须考虑共度较大的边中，实际上有多少条边在过程中被选中。

设 $H^*(G_i) := \bigcup_{k=k_0}^{k_1} H_k(G_i)$，注意，如果 e_i 是坏的且 $d_{e_i}(G_{i-1}) \in [K_\triangle/6, K_\triangle \ell)$，则 $e_i \in H^*(G_{i-1})$。设 $a^*(G_m)$ 为过程中所选边 e_i 的数量，使得 $e_i \in H^*(G_{i-1})$。由于

$$\begin{aligned}\sum_{i=1}^{m} d_{e_i}(G_{i-1})1_{\{e_i \text{ is bad}, d_{e_i}(G_{i-1})\in[K_\triangle/6, K_\triangle \ell)\}} &\leqslant K_\triangle \ell a^*(G_m) \\ &\leqslant \frac{2^{16} M(b,t)\ell a^*(G_m)}{b}\,,\end{aligned}$$

只需证明，对于某个常数 C，

$$\mathbb{P}\left(a^*(G_m) > Cb/\ell\right) \leqslant 2\exp(-4b)\,.$$

令 E 为如下事件:

$$h_k(G_i) \leqslant \frac{C_1 b^2}{k^2 2^{2k} t^4 n^2} \qquad \text{for all } k_0 \leqslant k \leqslant k_1 \text{ and } i \leqslant m\,.$$

根据命题 5.1，$\mathbb{P}(E^c) \leqslant \exp(-4b)$。现在我们证明，如果 C 足够大，那么 $\mathbb{P}(a^*(G_m) > Cb/\ell \text{ and } E) \leqslant \exp(-4b)$。这就完成了引理的证明。

在事件 E 发生时，每一步都有

$$\begin{aligned}|H^*(G_{i-1})| &\leqslant \sum_{k=k_0}^{k_1} \frac{C_1 b^2}{k^2 2^{2k} t^4 n^2} \\ &\leqslant \frac{2C_1 b^2}{k_0^2 2^{2k_0} t^4 n^2} \\ &\leqslant \frac{C_2 b^2}{\ell^2 K_\triangle^2} \\ &\leqslant \frac{C_2 b^4}{\ell^2 M(b,t)^2}\,, \end{aligned} \tag{6.23}$$

对于某个常数 C_2，其中我们用到了 (6.21) 中的不等式 $k_0 \geqslant \ell/2$ 。因此，事件“$a^*(G_m) > Cb/\ell$ 且 E”意味着

(i) 每一步中，$H^*(G_{i-1})$ 内至多有 $C_2 b^4/\ell^2 M(b,t)^2$ 个点对；(ii) 至少有 Cb/ℓ 步选择了 $e_i \in H^*(G_{i-1})$。

由 (i) 可知，给定 G_{i-1} 时，在 $H^*(G_{i-1})$ 中选中 e_i 的条件概率始终至多为

$$\frac{C_2 b^4}{\ell^2 M(b,t)^2 (N-i+1)} \leqslant \frac{2C_2 b^4}{\ell^2 M(b,t)^2 N}\,.$$

因此，对于任意序列 $i_1 < i_2 < \cdots < i_j$，给定 $j = Cb/\ell$，在所有这些步骤中于 $H^*(G_{i-1})$ 选中 e_i 的概率至多为

$$\left(\frac{2C_2 b^4}{\ell^2 M(b,t)^2 N}\right)^j\,.$$

对至多 $\binom{m}{j} \leqslant (em/j)^j$ 种序列选择应用并集界，可得

$$\mathbb{P}\left(a^*(G_m) > Cb/\ell \,\text{and}\, E\right) \leqslant \left(\frac{2emC_2 b^4}{j\ell^2 M(b,t)^2 N}\right)^j\,.$$

我们现在利用 $m/N = t$、$M(b,t)^2 \geqslant$ 团$(b,t)^2 = b^3/\ell^3$ 以及 $j = Cb/\ell$ 的值来推导得出】【。

$$\mathbb{P}\left(a^*(G_m) > Cb/\ell \,\text{and}\, E\right) \leqslant \left(\frac{2eC_2 t\ell^2}{C}\right)^{Cb/\ell}\,.$$

如果 C 取得足够大，那么方括号中的表达式至多为 $t^{1/2}$ =exp$(-\ell/2)$，因此

$$\mathbb{P}\left(a^*(G_m) > Cb/\ell \,\text{and}\, E\right) \leqslant \exp(-Cb/2) \leqslant \exp(-4b)\,.$$

□

7. Triangle counts in $G(n,m)$ – lower bounds on deviations

现在我们证明定理 1.1 的下界部分。也就是说，在四个区间中，我们证明相应量级的偏差（例如，在正态区间中至少为 $cN(b,t)$）发生的概率至少为 $\exp(-b)$。或许令人惊讶的是，这在正态区间中最具挑战性。

在正态区间（第 7.1 节）中，我们使用逆向 Freedman 不等式（引理 2.3）证明这一结果。在其他三个区间（星形图、枢纽和团）中，我们给出一种显式结构（图），它以至少 $\exp(-b)$ 的概率出现在 $G(n,m)$ 中，并直接导致三角形计数偏差。具体细节见第 7.2 节。

7.1. 正态区间。在本小节中，我们使用逆向 Freedman 不等式（引理 2.3）证明，对于某个绝对常数 $c>0$，有

$$\mathbb{P}\left(D_{\triangle}(G_m) \geqslant cb^{1/2}t^{3/2}n^{3/2}\right) \geqslant \exp(-b) \tag{7.1}$$

对于正态区间内的所有点对 (b,t)。该区间在图 2 中以绿色显示。若固定 $t \geqslant Cn^{-1/2}(\log n)^{1/2}$ 的值，则此范围对应于满足以下条件的 b 的取值：

$$3\log n \leqslant b \leqslant \min\{t^{1/3}n\ell^{4/3}\,,\, t^{3/2}n^{3/2}\ell^{3/2}\}\,.$$

我们还指出，在此区间内，截断值为

$$K_{\triangle} = 2^{16}b^{-1/2}t^{3/2}n^{3/2} \quad \text{and} K_{\wedge} = 2^{8}b^{-1/2}t^{1/2}n^{3/2}\,.$$

在上一节中，我们考虑了一个增量经过截断的过程，其最终值为 $D'_{\triangle}(G_m)$，该过程在式 (6.4) 中引入。回顾一下，

$$D'_{\triangle}(G_m) = \sum_{i=1}^{m}\left[3\,\frac{(N-m)_2(m-i)}{(N-i)_3}\,X'_{\wedge}(G_i) + \frac{(N-m)_3}{(N-i)_3}\,X'_{\triangle}(G_i)\right]$$

使用截断后的随机变量 $X'_{\wedge}(G_i) = X_{\wedge}(G_i)1_{X_{\wedge}(G_i)\leqslant K_{\wedge}}$ 和 $X'_{\triangle}(G_i) = X_{\triangle}(G_i)1_{X_{\triangle}(G_i)\leqslant K_{\triangle}}$，其中 $K_{\wedge} = 2^8M(b,t)/bt$ 和 $K_{\triangle} = 2^{16}M(b,t)/b$。

遗憾的是，截断使该过程成为超鞅而非鞅，因此我们无法应用 Freedman 逆向不等式（引理 2.3）。这让我们面临一个棘手的抉择，因为我们仍然希望对增量的大小给出一个界。

因此，我们将“修复”鞅性质。定义 $D''_{\triangle}(G_m) = \sum_{i=1}^{m}\mathbb{X}''_i$ 为重新平衡的增量之和

$$\mathbb{X}''_i := 3\frac{(N-m)_2(m-i)}{(N-i)_3}\,X''_{\wedge}(G_i) + \frac{(N-m)_3}{(N-i)_3}\,X''_{\triangle}(G_i)\,,$$

其中 $X''_{\wedge}(G_i)$ 和 $X''_{\triangle}(G_i)$ 定义为

$$X''_{\wedge}(G_i) = X'_{\wedge}(G_i) + \mathbb{E}\left[X_{\wedge}(G_i)1_{X_{\wedge}(G_i)>K_{\wedge}}|G_{i-1}\right]$$

and

$$X''_{\triangle}(G_i) \,=\, X'_{\triangle}(G_i) \,+\, \mathbb{E}\left[X_{\triangle}(G_i)1_{X_{\triangle}(G_i)>K_{\triangle}}|G_{i-1}\right] .$$

注意，这些随机变量已重新平衡，使得 $\mathbb{E}\left[\mathbb{X}''_i|G_{i-1}\right] \,=\, 0$。也就是说， $D''_{\triangle}(G_m)$ $=\sum_{i=1}^{m}\mathbb{X}''_i$是 $D''_{\triangle}(G_m)$ 的鞅表示。现在我们来给出 $D''_{\triangle}(G_m)$ 的一个下界。

命题 7.1. 存在一个绝对常数 $c>0$ ，使得下述结论成立。设 $n^{-1/2}(\log\, n)^{1/2} \leqslant t \leqslant 1/2$, 并假设 $3\log\, n \,\leqslant\, b \,\leqslant \min\{t^{1/3}n\ell^{4/3}\,,\, t^{3/2}n^{3/2}\ell^{3/2}\}$。则

$$\mathbb{P}\left(D''_{\triangle}(G_m) \,\geqslant\, cb^{1/2}t^{3/2}n^{3/2}\right) \,\geqslant\, \exp(-b)\,.$$

我们还证明了以下引理，它表明 $D''_{\triangle}(G_m)$ 很可能与 $D_{\triangle}(G_m)$非常接近。

引理 7.2. 存在一个绝对常数 $c_0 \,>\, 0$ ，使得对所有 $c \,<\, c_0$ ，下述结论均成立。设 $n^{-1/2}(\log\, n)^{1/2} \,\leqslant\, t \,\leqslant\, 1/2$, 并假设 $3\log\, n \,\leqslant\, b \,\leqslant c^3\min\{t^{1/3}n\ell^{4/3}\,,\, t^{3/2}n^{3/2}\ell^{3/2}\}$。则

$$\mathbb{P}\left(D''_{\triangle}(G_m) - D_{\triangle}(G_m) \,\geqslant\, \frac{cb^{1/2}t^{3/2}n^{3/2}}{2}\right) \,\leqslant\, \exp(-2b)\,,$$

对于所有充分大的 n。

显然，正态区间的下界 (7.1) 可由命题 7.1 和引理 7.2 推出。现在我们转而证明这两个结果。我们先从引理 7.2 开始。

我们继续使用记号 k_0 表示满足 $2^k t^2 n \leqslant K_{\triangle}/6$ 的最大整数 k 。还可以验证 $2^{k_0-8}tn \,\leqslant\, K_{\wedge}$。

证明。首先，我们回顾一下， $D'_{\triangle}(G_m) \,\leqslant\, D_{\triangle}(G_m)$ 是确定成立的（因为只有正增量会被截断），并且

$$D''_{\triangle}(G_m) \,-\, D'_{\triangle}(G_m) \,\leqslant\, \sum_{i=1}^{m} 3t\mathbb{E}\left[X_{\wedge}(G_i)1_{X_{\wedge}(G_i)>K_{\wedge}} \mid G_{i-1}\right] \,+\, \mathbb{E}\left[X_{\triangle}(G_i)1_{X_{\triangle}(G_i)>K_{\triangle}} \mid G_{i-1}\right] . \tag{7.2}$$

证明将基于以下事实：如果 C 是一个足够大的常数，那么每个事件的

$$\exists i \leqslant m \quad \mathbb{E}\left[X_{\wedge}(G_i)1_{X_{\wedge}(G_i)>K_{\wedge}}|G_{i-1}\right] \,>\, \frac{Cb}{n} \tag{7.3}$$

and

$$\exists i \leqslant m \quad \mathbb{E}\left[X_{\triangle}(G_i)1_{X_{\triangle}(G_i)>K_{\triangle}}|G_{i-1}\right] \,>\, \frac{Cbt}{n} \,+\, \frac{Cb^{5/2}}{t^{3/2}n^{7/2}\ell^2} \tag{7.4}$$

概率至多为 2 exp(−3b)。下面说明这已经足够。假设（7.3）和（7.4）中的事件均不发生。那么，由于求和涉及 $m \leqslant tn^2$ 项，（7.2）的右侧至多为

$$3Cbt^2n \,+\, Cbt^2n \,+\, \frac{Cb^{5/2}t^{1/2}}{n^{3/2}\ell^2}\,.$$

现在我们使用 $bt \leqslant c^3t^{4/3}n\ell^{4/3} \leqslant c^3n$ 和 $b^2 \leqslant c^6t^3n^3\ell^3 \leqslant c^6tn^3\ell^2$，并注意到这至多为

$$\big(4Cc^{3/2} \,+\, Cc^6\big)b^{1/2}t^{3/2}n^{3/2} \,\leqslant\, cb^{1/2}t^{3/2}n^{3/2}/2$$

只要 c 足够小，即可满足所需条件。

我们从 (7.3) 开始。我们指出，对于所有 $i \leqslant m$is $\Delta(G_m) \leqslant 64tn$，这里的期望值为 0；除概率至多为 exp(−96tn) 的情形外，这一结论都成立，见 (6.14)。现在假设 $b \geqslant 32tn$。这一论证在某些方面与第 6.2.1 节中的论证类似。同样，我们利用 $X_\wedge(G_i) \,\leqslant\, A_\wedge(G_i) \,=\, 4\mathsf{max}_{v\in e_i}\, d_v(G_{i-1})$，并且条件 $X_\wedge(G_i) > K_\wedge$ 意味着该最大值至少为 $K_\wedge/4 > 2^{k_0-8}tn$。现在只需注意，每个顶点被选作下一条边的一部分的概率至多为 $4/n$，因此

$$\begin{aligned}\mathbb{E}\Big[X_\wedge(G_i)1_{X_\wedge(G_i)>K_\wedge} \,\mid\, G_{i-1}\Big] \,&\leqslant\, \frac{16}{n}\sum_u d_u(G_{i-1})1_{d_u(G_{i-1})>2^{k_0-8}tn}\\ &\leqslant\, \frac{16}{n}\sum_{j=k_0-8}^{\log_2(1/t)} 2^{j+1}tn\,|V_j|\,.\end{aligned}$$

现在，根据引理 3.1，除概率至多为 exp(−3b) 的情况外，我们有 $|V_j| \leqslant b/tnj2^{j-8}$ 对所有 $j \geqslant 5$。在这种情况下，我们有

$$\begin{aligned}\mathbb{E}\Big[X_\wedge(G_i)1_{X_\wedge(G_i)>K_\wedge} \,\mid\, G_{i-1}\Big] \,&\leqslant\, \frac{2^{13}}{n}\sum_{j=k_0-8}^{\log_2(1/t)}\frac{b}{j}\\ &\leqslant\, \frac{2^{15}b}{n} \qquad (7.5)\end{aligned}$$

对所有 $i \leqslant m$，其中最后一个不等式利用了 (6.21) 中的 $k_0 - 8 \geqslant \ell/2$，以及求和中至多有 2ℓ 项。这就完成了对 (7.3) 的概率界估计。

现在我们转向（7.4）。如第 6.2.2 节所述，我们利用 that $X_\triangle(G_i) \,\leqslant\, 6d_{e_i}(G_{i-1})$，并根据每条边是良好边还是不良边来拆分各项。也就是说，我们将利用 that

$$\begin{aligned}&\mathbb{E}\big[X_\triangle(G_i)1_{X_\triangle(G_i)>K_\triangle}|G_{i-1}\big]\\ &\quad\leqslant\, \mathbb{E}\big[X_\triangle(G_i)1_{e_i\,\mathrm{good}}1_{X_\triangle(G_i)>K_\triangle}|G_{i-1}\big] \,+\, \mathbb{E}\big[X_\triangle(G_i)1_{e_i\,\mathrm{bad}}1_{X_\triangle(G_i)>K_\triangle}|G_{i-1}\big]\,.\end{aligned}$$

对于好边，论证与上文给出的基本相同。我们首先回顾引理 6.3 的证明，见（6.17）：对于好 e_i，我们有 $d_{e_i}(G_{i-1})$ 6

$128t\sum_{v\in e_i} d_v(G_m)1_{d_v(G_m)>K_\triangle/2^{10}t}$。利用每个顶点被包含在下一条边中的概率至多为$4/n$，我们得到

$$\begin{aligned}\mathbb{E}\left[X_\triangle(G_i)1_{e_i \text{ good}}1_{X_\triangle(G_i)>K_\triangle}|G_{i-1}\right] &\leqslant \frac{2^9 t}{n}\sum_v d_v(G_m)1_{d_v(G_m)>K_\triangle/2^{10}t}\\ &\leqslant \frac{2^9 t}{n}\sum_{j\geqslant k_0-8} 2^{j+1}tn|V_j|\,.\end{aligned}$$

我们现在可以完全按照上文（7.5）中的论证，得到

$$\mathbb{E}\left[X_\triangle(G_i)1_{e_i \text{ good}}1_{X_\triangle(G_i)>K_\triangle}|G_{i-1}\right] \leqslant \frac{2^{20}bt}{n}\,.$$

对于坏边，我们回顾 $X_\triangle(G_i)\leqslant 6d_{e_i}(G_{i-1})$。我们采用与 (6.22) 类似的论证，但用 k_0-3 替代 k_1。我们得到，除概率至多为 $\exp(-3b)$ 的情况外，有

$$\begin{aligned}\sum_{e\in E(K_n)\setminus E(G_{i-1})} d_e(G_{i-1})1_{\{e \text{ is bad},d_e(G_{i-1})>K_\triangle/6\}} &\leqslant \sum_{k=k_0-3}^{2\log_2(1/t)} 2^{k+1}t^2 n\frac{C_1b^2}{k^2 2^{2k}t^4n^2}\\ &\leqslant \sum_{k=k_0-3}^{2\log_2(1/t)}\frac{2C_1b^2}{k^2 2^k t^2 n}\\ &\leqslant \frac{4C_1b^2}{(k_0-3)^2 2^{k_0-3}t^2n}\\ &\leqslant \frac{32C_1b^2}{(k_0-3)^2K_\triangle}\\ &\leqslant \frac{2^{23}C_1b^3}{\ell^2 M(b,t)}\\ &= \frac{2^{23}C_1b^{3/2}}{t^{3/2}n^{3/2}\ell^2}\,,\end{aligned}$$

其中我们用到了 $k_0-3\geqslant \ell/2$，参见（6.21），以及在此范围内 $M(b,t)=b^{1/2}t^{3/2}n^{3/2}$。由于每一对作为下一条边被纳入的概率至多为 $4/n^2$，我们得到。

$$\mathbb{E}\left[X_\triangle(G_i)1_{\{e_i \text{ is bad}\}}1_{X_\triangle(G_i)>K_\triangle}|G_{i-1}\right] \leqslant \frac{Cb^{3/2}}{t^{3/2}n^{7/2}\ell^2}\,,$$

对于常数 $C=2^{23}C_1$。这完成了对（7.4）中概率的界定，也就完成了证明。□

现在我们转向命题 7.1 的证明。如上所述，我们的思路是使用 Freedman 下界不等式。为此，我们需要过程方差的一些下界。

在引理 6.2 中，对于 $b=n$，我们看到，除概率至多为 e^{-n} 的情形外，i 次增量的方差上界为 Ct^2n。下面的

结论给出了同阶的下界。这并不令人意外：由于共度通常以 $tn^{1/2}$ 的量级变化，人们会预期增量通常也处于这一量级。

引理 7.3. 存在一个绝对常数 $c_1 > 0$，使得以下结论成立。设 $n^{-1/2}/c_1 \leqslant t \leqslant 1/2$。对于所有 $m/2 \leqslant i \leqslant m$，有

$$\mathbb{P}\left(\mathbb{E}\left[(\mathbb{X}_i'')^2 | G_{i-1}\right] \leqslant c_1 t^2 n\right) \leqslant \exp(-c_1 n)$$

对于所有充分大的 n。

由于该引理的证明篇幅较长但相对初等，因此收录于附录 C。现在我们完成命题 7.1 的证明。

命题 7.1 的证明。固定 t 的取值，使其处于范围 $n^{1/2}(\log n)^{1/2} \leqslant t \leqslant 1/2$ 内，并固定 $3\log n \leqslant b \leqslant \min\{t^{1/3}n\ell^{4/3}, t^{3/2}n^{3/2}\ell^{3/2}\}$。事实上，我们可以假设 b 至多为 $\min\{t^{1/3}n\ell^{4/3}, t^{3/2}n^{3/2}\ell^{3/2}\}$ 的一个较小常数倍；否则，只需采用由该 b 值给出的下界，并相应调整常数。特别地，我们假设 $b \leqslant c^3\min\{t^{1/3}n\ell^{4/3}, t^{3/2}n^{3/2}\ell^{3/2}\} \leqslant c^2 n$，其中 $c \leqslant 2^{-36}$ 是引理 7.2 和 7.3 中常数的最小值。我们将证明概率的一个下界

$$\mathbb{P}\left(D_{\triangle}''(G_m) \geqslant cb^{1/2}t^{3/2}n^{3/2}\right)$$

使用 Converse Freedman 不等式（引理 2.3）。在整个证明过程中，我们令 $\alpha' = cb^{1/2}t^{3/2}n^{3/2}$。

这里有两个小问题。其一，引理 2.3 给出的是鞅在某个时刻 m 之前超过某个值 α' say 的概率下界，而不是在第 m 步超过它。解决办法很简单：改为证明我们有很大的概率到达 $2\alpha'$，然后利用我们的上界结果来控制过程随后下降的概率。[20] 中也考虑过同样的问题。

另一个问题是，我们无法像使用引理 2.3 所要求的那样，对 $\|\mathbb{X}_i''\|_\infty$ 给出足够好的界。因此，我们转而考虑该过程的一个变体，它极有可能与原过程相等，但在某个罕见事件发生时可能表现不同。令 E_i 表示对于某个特定的 i，事件 (7.3) 或 (7.4) 之一成立的事件。令E_{i-1}^+ 表示存在一对 $e \in E(K_n) \setminus E(G_{i-1})$，使得如果 $e_i = e$，那么 E_i 将会发生的事件。再令 $E^+ = \bigcup_{i=1}^m E_{i-1}^+$。由于每一对 $e \in E(K_n) \setminus E(G_{i-1})$ 的概率至少为 n^{-2}，我们得到】【。

$$\mathbb{P}\left(E^+\right) \leqslant \sum_{i=1}^m \mathbb{P}\left(E_{i-1}^+\right) \leqslant n^2 \sum_{i=1}^m \mathbb{P}\left(E_i\right) \leqslant 4n^2 \exp(-3b), \tag{7.6}$$

其中，最后一个不等式依赖于引理 7.2 的证明中给出的事件 (7.3) 和 (7.4) 的概率界。令 $X_i^*$by

$$X_i^* := \mathbb{X}_i'' 1_{(E_{i-1}^+)^c}.$$

由于该事件是 G_{i-1}-可测的，因此我们有 X_i^* 仍然表现为鞅增量，其意义在于 $\mathbb{E}\left[X_i^* \mid G_{i-1}\right] = 0$。令 $D_{\triangle}^*(G_m) = \sum_{i=1}^m X_i^*$。我们观察到

$$\mathbb{P}\left(D_{\triangle}''(G_m) \neq D_{\triangle}^*(G_m)\right) \leqslant 4n^2 \exp(-3b) \leqslant \exp(-2b)\,.$$

此外， X_i^* 只有在 E_{i-1}^+ 不发生时才非零；此时，(7.3) 和 (7.4) 中的期望值分别至多为 $C_1 b/n$ 和 $C_1 bt/n + C_1 b^{5/2}/t^{3/2}n^{7/2}\ell^2$ ，其中 C_1为某个常数。在这一事件下，我们有

$$|\mathbb{X}_i'' - \mathbb{X}_i'| \leqslant \frac{C_2 bt}{n} + \frac{C_2 b^{5/2}}{t^{3/2}n^{7/2}\ell^2} \leqslant \frac{C_3 \mathrm{NORMAL}(b,t)}{b}$$

for 常数 C_2、 C_3，其中最后一个界由 $b^3 \leqslant tn^3\ell^4 \leqslant t^3n^5\ell^2 \leqslant tn^5$推出。一个 s

$$|\mathbb{X}_i'| \leqslant 3t|X_{\wedge}'(G_i)| + |X_{\triangle}'(G_i)| \leqslant \frac{C_4 M(b,t)}{b} = \frac{C_4 \mathrm{NORMAL}(b,t)}{b}$$

因此，在此情形下，我们有一个绝对界

$$\begin{aligned} |X_i^*| &\leqslant |\mathbb{X}_i''| 1_{(E_{i-1}^+)^c} \\ &\leqslant |\mathbb{X}_i'| + |\mathbb{X}_i'' - \mathbb{X}_i'| 1_{(E_{i-1}^+)^c} \\ &\leqslant \frac{C \mathrm{NORMAL}(b,t)}{b}\,, \end{aligned}$$

其中 C为某个常数。

根据引理 7.3，除了概率至多为 n^2exp(−cn)的情况外，所有增量都满足

$$\mathbb{E}\left[(\mathbb{X}_i'')^2 \,|\, G_{i-1}\right] \geqslant ct^2 n\,.$$

It f由此可知，除非概率至多为 n^2exp(−cn) 6 exp(−2b)，否则我们有

$$\sum_{i=1}^m \mathbb{E}\left[(\mathbb{X}_i'')^2 \,|\, G_{i-1}\right] \geqslant ct^3 n^3/10\,.$$

由于对于任意 i， $X_i^* \neq \mathbb{X}_i''$ 发生的概率至多为 exp(−2b)，因此我们有

$$V(m) := \sum_{i=1}^m \mathbb{E}\left[(X_i^*)^2 \,|\, G_{i-1}\right] \geqslant ct^3 n^3/10\,,$$

除非概率至多为 exp(−b)。

我们将对 $\alpha = 2\alpha' = 2cb^{1/2}t^{3/2}n^{3/2}$、 $\beta = ct^3n^3/10$ 和$R = 2$NORMAL(b, t) $/b = b^{-1/2}t^{3/2}n^{3/2}$应用引理 2.3。我们可以取 $\delta = 1$。于是得到

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^{m'} X_i^* \geqslant \alpha \text{ for some } m' \leqslant m\right) &\geqslant \frac{1}{2}\exp\left(\frac{-3\alpha^2}{\beta}\right) - \exp(-b) \\ &\geqslant \exp(-120cb) - \exp(-b) \\ &\geqslant \exp(-b/2) - \exp(-b) \\ &\geqslant 4\exp(-b)\,. \end{aligned}$$

令 τ 为与事件 $\sum_{i=1}^{m'} X_i^* \geqslant \alpha$相关的命中时刻，即使其发生的最小 m'。现在我们证明

$$\mathbb{P}\left(D_{\triangle}^*(G_m) \geqslant \alpha' \mid \tau \leqslant m\right) \geqslant \frac{1}{2}.$$

根据标准论证，只需证明对于任意固定值 $\tau = m' < m$，条件概率至少为$1/2$ 。现在，在此事件上，实际上极有可能 $D_{\triangle}^*(G_m) = \sum_{i=1}^{m} X_i^* \geqslant \alpha' = \alpha/2$。这不成立的条件概率至多为

$$\mathbb{P}\left(\left|\sum_{i=m'+1}^{m} X_i^*\right| \geqslant \alpha' \mid \tau = m'\right)$$

容易看出，通过采用第6节中的论证方式，其至多为 $\exp(-cb) \leqslant 1/2$ 。这表明，$D_{\triangle}^*(G_m) \geqslant \alpha$ 的概率至少为 $2\exp(-b)$。最后，由于它们以至多 $\exp(-2b)$ 6 $\exp(-b)$ 的概率不同，我们有 $D_{\triangle}''(G_m) \geqslant \alpha$ 的概率至少为 $\exp(-b)$。□

7.2. 其余区间的下界。我们回顾一下，我们必须证明在这些区间中，对于某个常数 $c > 0$，有

$$N_{\triangle}(G_m) > \mathbb{E}\left[N_{\triangle}(G_m)\right] + cM(b,t)$$

概率至多为 $\exp(-b)$。在这些区间中，我们总有 $b \geqslant 6\log n$。

在其余三个区间中，我们给出一个图 G_*，它至多包含 $b/10\ell$条边，并且满足

$$\mathbb{E}\left[N_{\triangle}(G_{m-b/10\ell} \cup G_*)\right] > \mathbb{E}\left[N_{\triangle}(G_m)\right] + 2cM(b,t). \tag{7.7}$$

注意，这意味着

$$\mathbb{E}\left[N_{\triangle}(G_m) | G_m \supseteq G_*\right] > \mathbb{E}\left[N_{\triangle}(G_m)\right] + 2cM(b,t),$$

因为我们可以将 G_m （以 $G_m \supseteq G_*$ 为条件）耦合为包含 $G_{m-b/10\ell} \cup G_*$。我们将看到，这足以证明下界。首先注意

$$\mathbb{P}\left(G_* \subseteq G_m\right) \geqslant \frac{(m)_{b/10\ell}}{(N)_{b/10\ell}} \geqslant (t/2)^{b/10\ell} \geqslant \exp(-b/2).$$

Also

$$\mathbb{E}\left[N_{\triangle}(G_m) \mid G_m \supseteq G_*\right] \leqslant$$

$$\mathbb{E}\left[N_{\triangle}(G_m)\right] + cM(b,t) + n^3\mathbb{P}\left(N_{\triangle}(G_m) > \mathbb{E}\left[N_{\triangle}(G_m)\right] + cM(b,t) \;\middle|\; G_m \supseteq G_*\right),$$

由此可得

$$\mathbb{P}\left(N_{\triangle}(G_m) > \mathbb{E}\left[N_{\triangle}(G_m)\right] + cM(b,t) \,|\, G_m \supseteq G_*\right) \geqslant n^{-3}$$

因此

$$\mathbb{P}\left(N_{\triangle}(G_m) > \mathbb{E}\left[N_{\triangle}(G_m)\right] + cM(b,t)\right) \geqslant n^{-3}\exp(-b/2) \geqslant \exp(-b),$$

正如所要求的。

现在我们介绍在三个区间中给出所需下界的三种结构（子图）。

星形区域：在星形区域中，我们可以取 G_* 为度为 $b/10\ell$ 的星形图。

回忆一下，$\mathsf{STAR}(b,t) := b^2t\ell^{-2}1_{b\leqslant n\ell}$。在 $G_{m-b/10\ell}\cup G_*$ 中，涉及星的中心处顶点 G_* 的三角形的期望数量约为 $b^2t\ell^{-2}$。由于一条典型的边约属于 t^2n 个三角形，$\mathbb{E}\left[N_\triangle(G_m)\right]-\mathbb{E}\left[N_\triangle(G_{m-b/10\ell})\right]$ 度为 $b/10\ell$ 的星形图的期望数量将

枢纽区域：在枢纽区域中，我们可以取 G_* 为一个由 $b/10n\ell$ 个满度顶点组成的枢纽，即这些顶点的度为 $n-1$。

团区域：在团区域中，我们将 G_* 设为一个具有 $b^{1/2}/4\ell^{1/2}$ 个顶点的团。

8. Lower bounds for cherries

现在我们来考察樱桃的相应下界，即定理 1.8 的下界部分。正如我们在引言中提到的，我们仅给出证明概要。

正如我们在第4节中指出的，对于每个固定的 $t\geqslant 2n^{-1}\log n$ 值，正态区间对应于 $3\log n\leqslant b\leqslant t^{2/3}n\ell^{4/3}$，星形区域对应于 $t^{2/3}n\ell^{4/3}<b\leqslant n\ell$，而枢纽区域对应于 $b>n\ell$。

星形图和枢纽区域的情况都很直接。在星形情形下，假设$t^{2/3}n\ell^{4/3}<b\leqslant n\ell$，我们必须证明偏差至少为 cb^2/ℓ^2 的概率不小于 $\exp(-b)$。图中的第一个顶点的度至少为 $b/2\ell$的概率不小于 $t^{(1+o(1))b/2\ell}\geqslant\exp(-b)$。在这种情况下，仅这个顶点就参与了 $\Omega(b^2/\ell^2)$ 个樱桃。我们可以像上一节处理三角形时那样，通过对相应事件进行条件化，使这一论证更加严谨。

在枢纽区域，假设 $b\geqslant n\ell$，我们必须证明偏差至少为 cbn/ℓ 的概率不小于 $\exp(-b)$。至少存在 $t^{(1+o(1))bn/2n\ell}\geqslant\exp(-b)$ 的概率，使得有 $b/2n\ell$ 个顶点的度为 $n-1$。仅这些顶点就构成了 $\Omega(bn/\ell)$ 个樱桃。同样，可以利用条件化使这一论证严格化。

在正态区间内，假设 $3\log n\leqslant b\leqslant t^{2/3}n\ell^{4/3}$，我们必须证明，大小至少为 $cb^{1/2}tn^{3/2}$ 的偏差发生的概率至少为 $\exp(-b)$。类比上一节，可以将 $D''_\wedge(G_m)$ 定义为经过“修正”的鞅增量 $X''_\wedge(G_i)$ 之和。容易证明，$|D_\wedge(G_m)-D''_\wedge(G_m)|=\Omega(b^{1/2}tn^{3/2})$ 发生的概率至多为 $\exp(-2b)$，因此，只需证明 $D''_\wedge(G_m)$ 的界即可。

通过与上一节类似的论证可以证明，所有度都集中在平均值附近，且变化不超过 $t^{1/2}n^{1/2}$ 的尺度，这种情况极不可能发生，其概率至多为 $\exp(-\Omega(n))$。这使我们能够证明，以极高的概率，在整个过程中 $\mathbb{E}\left[X_\wedge(G_i)^2 \mid G_{i-1}\right] \geqslant c'tn$，修复过程也同样如此，$\mathbb{E}\left[X''_\wedge(G_i)^2 \mid G_{i-1}\right] \geqslant c'tn$。所需的下界现在可由 Converse Freedman 不等式（引理 2.3）结合 $R = 2^{10}tn$ 和 $\beta = c't^2n^3/2$ 得出，[2]，与命题 7.1 的证明类似。

9. Triangles in $G(n,p)$

在这里，我们推导出 $G(n,p)$ 中的结果，即定理 1.5 和定理 1.7。

我们从 $G(n,p)$ 的结果中推导出 $G(n,m)$ 的结果，这与 [20]及其他文章中的做法相同。具体而言，我们结合使用定理 1.1（该定理控制 $N_\triangle(G_m)$ 的偏差概率）以及表达式[3]

$$\mathbb{P}\left(N_\triangle(G_p) > \delta_n p^3(n)_3\right) = \sum_{m=0}^{N} b_N(m)\,\mathbb{P}\left(N_\triangle(G_m) > (1+\delta_n)p^3(n)_3\right), \qquad (9.1)$$

其中 $G_p \sim G(n,p)$ 和 $b_N(m) := \mathbb{P}(Bin(N,p) = m)$。

因此，掌握以下关于 $b_N(k)$ 的估计是有用的，而且

$$B_N(k) := \mathbb{P}\left(\mathrm{Bin}(N,p) \geqslant k\right).$$

以 $x_N = \frac{k-pN}{\sqrt{pqN}}$ 表述的结果，对于 $p \in (0,1)$ 是常数或 $p = p_N$ 函数的情形均成立。该结果的这一版本陈述于 [20], ，它可由 Bahadur [3] 的定理 2 轻松推出。其中包含表达式

$$E(x,N) := \sum_{i=1}^{\infty} \frac{(p^{i+1} + (-1)^i q^{i+1})x^{i+2}}{(i+1)(i+2)p^{i/2}q^{i/2}N^{i/2}},$$

其中 q 表示 $1-p$，后续陈述中也将如此。

Th定理 9.1。设（x_N）是一个满足 $1 \ll x_N \ll \sqrt{pqN}$ 的序列。 n

$$b_N(\lfloor pN + x_N\sqrt{pqN}\rfloor) = (1+o(1))\frac{1}{\sqrt{2\pi pqN}}\exp\left(-\frac{x_N^2}{2} - E(x_N,N)\right)$$

and

$$B_N(pN + x_N\sqrt{pqN}) = (1+o(1))\frac{1}{x_N\sqrt{2\pi}}\exp\left(-\frac{x_N^2}{2} - E(x_N,N)\right).$$

[2] 从技术上说，可能有必要引入一个 X^* 过程，就像我们在命题 7.1 的证明中所做的那样。[3] 请注意，只需考虑对 $G(n,\,p)$ 中边的数量进行条件化，即可轻松证明 (9.1)。

本节证明定理 1.5 和定理 1.7。我们从定理 1.5 开始，该定理给出了在参数 δ、p 的某个范围内，$r(\delta_n, p, n)$ 的渐近值。我们首先陈述这一结果的一个更精确版本。定义

$$\tilde{M}(\delta, p) := CM(\delta^2 pn^2, 2p)$$

其中，M 的定义见引言中的式 (1.1)，而 C 是定理 1.1 中的常数。我们还定义了

$$x_* := \left[(1+\delta_n)^{1/3} - 1\right] \sqrt{\frac{pN}{q}}.$$

命题 9.2。设 $n^{-1/2}(\log n)^{1/2} \ll p \ll 1$，且 δ_n 是满足条件的序列。

$$p^{-1/2}n^{-1} \ll \delta_n \ll p^{3/4}(\log n)^{3/4}, n^{-1/3}(\log n)^{2/3} + p\log(1/p).$$

Then

$$r(\delta_n, p, n) = \frac{x_*^2}{2} + E(x_*, N) + \log x_* + O\left(\frac{\delta_n \tilde{M}(\delta_n, p)}{p^2 n}\right) + O(1).$$

很容易看出，命题 9.2 蕴含定理 1.5。首先注意到 $x_* = (1 + o(1))\delta_n p^{1/2} n/3\sqrt{2}$，由此可得 $x_* \ll \sqrt{pqN}$ from which is follows that$E(x_*, N) \ll x_*^2$。现在我们将检验 that

$$\tilde{M}(\delta_n, p) \ll \delta_n p^3 n^3 \tag{9.2}$$

在 δ_n 给定的范围内。为此，我们检查用于定义$M(b, t)$ 的四个函数。注意，对所有 δ，NORMAL$(\delta^2 pn^2, 2p) \ll \delta p^3 n^3$，并且当 $\delta \ll p^{3/4}(\log n)^{3/4}$ 时，CLIQUE$(\delta^2 pn^2, 2p) \ll \delta p^3 n^3$ 。对于星形图和枢纽区域，相应的上界分别为 $\delta \ll n^{-1/3}(\log n)^{2/3}$ 和 $\delta \ll p\log(1/p)$。不过，我们注意到这些区间的边界出现在 $\delta^2 pn^2 = n\ell$ 时，也就是$\delta = p^{-1/2}n^{-1/2}\ell^{1/2}$，而相应区间所需的上界是这两个函数中的最大值。因此，界$\delta \ll n^{-1/3}(\log n)^{2/3} + p\log(1/p)$ 已足够。

从应用角度来看，式 (9.2) 背后的直观含义是，$G(n, m)$ 对偏差的贡献很小。事实上，量 $M(\delta_n, p)$ 可作为这一 $G(n, m)$ 贡献的上界。确实，由定理 1.1 可得

$$\mathbb{P}\left(N_\triangle(G_m) > \mathbb{E}\left[N_\triangle(G_m)\right] + \tilde{M}(\delta_n, p)\right) \leqslant \exp(-\delta^2 pn^2) \tag{9.3}$$

f对所有 $m \leqslant 2pN$ 均成立。由于 $\delta_n^2 pn^2 \geqslant x_*^2$，在此情形下，这一概率可以忽略不计 .

现在我们证明命题 9.2。这需要我们证明概率 $\mathbb{P}(N_\triangle(G_p) \geqslant (1+\delta_n)p^3(n)_3)$ 的上界和下界。我们定义

$$m_* := pN(1+\delta_n)^{1/3}$$

即（近似） m 的值，在该值处 $\mathbb{E}\,[N_{\triangle}(G_m)] = (1+\delta_n)p^3(n)_3$。换言之，如果 G_p 至少有这么多边，那么为了达到所述偏差，就不需要 $G(n,m)$ 偏差。考虑以下内容也会很有用

$$m_- \;:=\; m_* \;-\; 2p^{-2}n^{-1}\tilde{M}(\delta_n,p)\,.$$

设 $x_- := (m_- - pN)/\sqrt{pqN}$ 为 x 的对应值。注意 $x_- = x_* - \Theta(\tilde{M}(\delta_n,p)/p^{5/2}n^2)$，且 $x_* = \Theta(\delta_n p^{1/2}n)$。

命题 9.2 的证明。 $\mathbb{P}(N_{\triangle}(G_p) \geqslant (1+\delta_n)p^3(n)_3)$ 的下界相对直接。利用 (9.1) 和单调性，我们有

$$\mathbb{P}\left(N_{\triangle}(G_p) \;\geqslant\; (1+\delta_n)p^3(n)_3\right) \;\geqslant\; B_N(m_*)\mathbb{P}\left(N_{\triangle}(G_{m_*}) \geqslant (1+\delta_n)p^3(n)_3\right)\,.$$

我们还注意到， $G(n,m_*)$ 中三角形数量的期望为 $(1+\delta_n)p^3(n)_3 + O(p^2n)$，因此它与 $(1+\delta_n)p^3(n)_3$ 之间的差远小于 $t^{3/2}n^{3/2}$； $t^{3/2}n^{3/2}$ 是该随机变量在 $G(n,m)$ 中标准差的量级。由中心极限定理可知，上述后一个概率为 $\frac{1}{2} + o(1) \;=\; \exp(O(1))$。由此可得，

$$\mathbb{P}\left(N_{\triangle}(G_p) \;\geqslant\; (1+\delta_n)p^3(n)_3\right) \;\geqslant\; \exp\left(\frac{-x_*^2}{2} \;-\; E(x_*,N) \;-\; \log x_* \;+\; O(1)\right)$$

这给出了所需的界，即速率 $r(\delta_n,p,n$ 的上界).

现在我们证明 $\mathbb{P}(N_{\triangle}(G_p) \geqslant (1+\delta_n)p^3(n)_3)$ 的上界。再次使用 (9.1)。这次我们只分别考虑高于 m_- 和低于 m_- 的部分。高于 m_- 的部分至多为

$$\begin{aligned}
B_N(m_-) \;&\leqslant\; \exp\left(\frac{-x_-^2}{2} \;-\; E(x_-,N) \;-\; \log x_- \;+\; O(1)\right)\\
&\leqslant\; \exp\left(\frac{-x_*^2}{2} \;+\; O\left(\frac{x_*\tilde{M}(\delta_n,p)}{p^{5/2}n^2}\right) \;-\; E(x_*,N) \;-\; \log x_* \;+\; O(1)\right)\\
&\leqslant\; \exp\left(\frac{-x_*^2}{2} \;+\; O\left(\frac{\delta_n\tilde{M}(\delta_n,p)}{p^2n}\right) \;-\; E(x_*,N) \;-\; \log x_* \;+\; O(1)\right)
\end{aligned}$$

如所需。

现在对于 $m \leqslant m_-$。我们断言，

$$\mathbb{P}\left(N_{\triangle}(G_{m_-}) \geqslant (1+\delta_n)p^3(n)_3\right) \;\leqslant\; \exp(-\delta_n^2 pN)\,. \tag{9.4}$$

由于该概率远小于所需的界，且由单调性可知此结论对 $m \leqslant m_-$成立，因此证明完毕。

为证明 (9.4)，我们注意到 $N_{\triangle}(G_{m_-}) \geqslant (1+\delta_n)p^3(n)_3$ 对应于 $G(n,m)$ 中三角形计数的较大正偏差。事实上

$$\begin{aligned}\mathbb{E}\left[N_{\triangle}(G_{m_-})\right] &\leqslant \left(\frac{m_-}{N}\right)^3 (n)_3 \\ &= \left(\frac{m_* - 2p^{-2}n^{-1}\tilde{M}(\delta_n,p)}{N}\right)^3 (n)_3 \\ &\leqslant \left(\frac{m_*}{N}\right)^3 (n)_3 - \tilde{M}(\delta_n,p) \\ &= (1+\delta_n)p^3(n)_3 - \tilde{M}(\delta_n,p)\,.\end{aligned}$$

因此我们看到，在 G_{m_-} 中有$(1+\delta_n)p^3(n)_3$ 个三角形，代表着一个较大的正偏差，至少为 $M(\delta_n,p)$。如上文(9.3)所述，这种情况的概率至多为 $\exp(-\delta_n^2 pN)$。至此，(9.4)的证明以及该命题的证明均告完成。□

现在我们证明局部化情形下的结果，即定理 1.7。我们回顾一下，该定理指出，对于这一情形，

$$p^{3/4}(\log n)^{3/4}\,,\, n^{-1/3}(\log n)^{2/3} + p\log(1/p) \leqslant \delta_n \leqslant 1$$

我们有

$$r(\delta_n,p,n) = \Theta(1)\min\{\delta_n^{2/3}p^2n^2\log n, \delta_n^{1/2}pn^{3/2}\log n + \delta_n p^2n^2\log(1/p)\}\,. \quad (9.5)$$

在这种情形下，实现所需偏差的最简单方式是通过 G_p 包含某种特定结构（星形图、枢纽或团）。这些例子将为我们提供下界。

对于上界，我们将证明，仅仅增加没有任何结构的边，并不能让结果偏离太多。令 $m_0 = pN + \delta pn^2/10$。注意到这一点会很有用，因为

$$B_N(m_0) \leqslant \exp\left(\frac{-\delta_n^2 pn^2}{200}\right)\,. \quad (9.6)$$

我们对一些（标准的）下界的表述略显不拘谨。

定理 1.7 的证明。对应于导致额外三角形的最可能原因（枢纽、星形图或团），这三个部分的证明将采用类似的方法。在枢纽区域中，

$$p\log(1/p)\,,\, \frac{1}{p^2n} \leqslant \delta_n \leqslant 1$$

(9.5) 中的最小值为 $\Theta(\delta_n p^2n^2\log(1/p))$。注意，在此区域（当 $\delta_n \geqslant p\log(1/p)$ 时），由 (9.6) 可得 $\mathbb{P}(e(G_p) > m_0) \leqslant \exp(-\Omega(\delta_n p^2n^2\log(1/p)))$。

现在可以直接给出该区域内概率$\mathbb{P}(N_{\triangle}(G_p) \geqslant (1+\delta_n)p^3(n)_3$的上界和下界。下界可通过要求某些 $\Theta(\delta_n p^2 n)$个顶点的度为 $n-1$来得到。其概率为 $\exp(-O(\delta_n p^2n^2\log(1/p)))$

并因此产生额外的 $\delta_n p^3 n^3$ 个三角形，正如所需。对于上界，我们注意到

$$\begin{aligned}
&\mathbb{P}\left(N_{\triangle}(G_p) \geqslant (1+\delta_n)p^3(n)_3\right) \\
&\quad \leqslant \mathbb{P}\left(e(G_p) > m_0\right) + \mathbb{P}\left(N_{\triangle}(G_p) \geqslant (1+\delta_n)p^3(n)_3 \mid e(G_p) \leqslant m_0\right) \\
&\quad \leqslant \exp(-\Omega(\delta_n p^2 n^2 \log(1/p))) + \mathbb{P}\left(N_{\triangle}(G_{m_0}) \geqslant (1+\delta_n)p^3(n)_3\right) .
\end{aligned}$$

因此，只需证明第二个概率满足相同的界。容易验证， $\mathbb{E}\left[N_{\triangle}(G_{m_0})\right] \leqslant (1+9\delta_n/10)p^3(n)_3$，因此事件 $N_{\triangle}(G_{m_0}) \geqslant (1+\delta_n)p^3(n)_3$ 表示 $N_{\triangle}(G_{m_0})$ 中的正偏差 $\Theta(\delta_n p^3 n^3)$。根据定理 1.1，该事件的概率至多为 $\exp(-\Omega(\delta_n p^2 n^2 \log(1/p)))$（在此区域，$\mathrm{HUB}(b,t)$ 是最大值点），证毕。

另外两个区域中的证明非常相似。在星形区域中，

$$\frac{1}{p^6 n^3} , n^{-1/3}(\log n)^{2/3} \leqslant \delta_n \leqslant \frac{1}{p^2 n}$$

(9.5) 中的最小值为 $\Theta(\delta_n^{1/2} p n^{3/2} \log n)$；类似地，在此区域（因为 $\delta_n \geqslant n^{-1/3}(\log n)^{2/3}$），由 (9.6) 可得 $\mathbb{P}(e(G_p) > m_0) \leqslant \exp(-\Omega(\delta_n^{1/2} p n^{3/2} \log n))$。

下界可通过要求某个顶点的度至少为 $\delta_n^{1/2} p n^{3/2}$ 来得到。该事件的概率为 $\exp(-O(\delta_n^{1/2} p n^{3/2} \log n))$[4]，并由此产生额外的 $\delta_n p^3 n^3$ 个三角形，正如所需。上界与枢纽情形完全相同，只是定理 1.1 中的极值者现在是 $\mathrm{STAR}(b,t)$，从而得到界 $\exp(-\Omega(\delta_n^{1/2} p n^{3/2} \log n))$。

在团区域中，

$$p^{3/4}(\log n)^{3/4} \leqslant \delta_n \leqslant \frac{1}{p^6 n^3}$$

(9.5) 中的最小值为 $\Theta(\delta_n^{2/3} p^2 n^2 \log n)$。由于 $\delta_n \geqslant p^{3/4}(\log n)^{3/4}$，由 (9.6) 可得 $\mathbb{P}(e(G_p) > m_0) \leqslant \exp(-\Omega(\delta_n^{2/3} p^2 n^2 \log n))$。下界来自在 $2\delta_n^{1/3} pn$ 个顶点上植入一个团，该团至少产生 $\delta_n p^3 n^3$ 个额外三角形，且其概率为 $\exp(-O(\delta_n^{2/3} p^2 n^2 \log n))$。上界与上述情况相同，只是现在 $\mathrm{CLIQUE}(b,t)$ 是定理 1.1 中的极值者；于是我们得到界 $\exp(-\Omega(\delta_n^{2/3} p^2 n^2 \log n))$，正如所需。□

10. Concluding Remarks

一个很自然的问题是：能否更精确地理解本文讨论的尾概率？例如，我们能否在一般情况下求出$G(n,m)$和 $G(n,p)$模型中偏差概率对数的渐近值?

在这两种模型中，这一目标在某些区域内均已实现。例如，在稠密情形下，[20]针对两种模型的正态区间都给出了此类结果。

[4]注意，在此情形下，$\log n$ 和$\log(1/p)$是等价的（同阶的）

这些结果还涵盖了略微稀疏一些的随机图中正态区间的一部分。并且，正如我们在引言中所讨论的，$G(n,p)$ 也为三角形计数的大偏差对应的对数概率找到了渐近紧界。一个值得探究的问题是，这些结果能否扩展到涵盖所有局部化区域（星形图、枢纽和团）。

我们还可以探讨表现出某些偏差的随机图的典型结构。稍作努力，似乎就能证明我们关于三角形的结果的弱结构性版本。例如，给定一个对应于团区域的点对 t, a，我们可以问：在条件为 G_m 包含的三角形数量比预期多 a 个的情况下，它是否很可能包含一个具有 Θ 个顶点的较稠密子图（$a^{1/3}$）？若能证明一个更强的结构性结果，指出在这种情况下 G_m 包含一个由 $(1+o(1))a^{1/3}$ 个顶点构成的团（或至少一个非常稠密的子图），将会非常有趣。关于 $G(n,p)$ 中大偏差的相关结果，参见 [21] 的定理 1.8。

自然也可以问，这些结果能否推广到其他图。这些结果是在 [20]，中针对相对稠密的随机图得到的，但若能了解这类结果的适用范围将会很有意思。

我们还要提及一个初等问题。这个问题似乎一般情况下尚未解决，并且与引理 5.5 有关。我们用 $\mathcal{G}_{n,m}$ 表示顶点集为 $[n]=\{1,\dots,n\}$ 且有 m 条边的所有图组成的集合。给定图 G 的一个顶点集 S，令 $N(S):=\bigcup_{v\in S}N(v)$ 表示顶点 $v\in S$ 的邻域之并。

问题 1。令 $2\leqslant k\leqslant n\leqslant m$。$\min_{G\in\mathcal{G}_{n,m}}\max_{S\subseteq[n]:|S|=k}|N(S)|$ 的值是多少？

特别地，可以考虑以下情形：k 是常数，n 很大，以及对于某个 $p\in(0,1)$，$m\approx p\binom{n}{2}$。两个值得考虑的自然候选极值图是：

(i) 具有 $\approx p^{1/2}n$ 个顶点的完全图，(ii) 密度为 p 的随机图（或准随机图）。

在情况 (i) 中，优化 $|N(S)|$ 是显而易见的，因为对于任何包含完全图中某个顶点的 S，都有 $|N(S)|\approx p^{1/2}n$。而在情况 (ii) 中，对于所有满足 $|S|=k$ 的子集 S，$|N(S)|$ 的大小大致相同；事实上，$|N(S)|\approx(1-(1-p)^k)n$。因此，我们尤其可以提出以下问题：

问题 2. 设 $k\geqslant 2$，并令 $p\in(0,1)$。当 $n\to\infty$ 时，在极限情况下，以下命题是否成立？

$$\min\left\{\max_{S\subseteq[n]:|S|=k}|N(S)|\,:\,G\in\mathcal{G}_{n,p\binom{n}{2}}\right\}=(1+o(1))\min\{p^{1/2}n,(1-(1-p)^k)n\}\,?$$

References

1. Fanny Augeri。关于 Wigner 矩阵和稀疏 Erd˝os–R´enyi 随机图应用的非线性大偏差界。《概率年刊》，48(5)：2404–2448，2020。

2. K. Azuma。某些相依随机变量的加权和。《东北数学杂志》，19(3)：357–367，1967。3. R. R. Bahadur。二项分布函数的若干近似。《数学统计年刊》，31(1)：43–54，1960。4. A. Barbour、M. Karo´nski 和 A. Ruci´nski。可分解随机变量的中心极限定理及其在随机图中的应用。《组合理论杂志》B辑，47：125–145，1989。5. B. Bollob´as。随机图的度序列。《离散数学》，33(1)：1–19，1981。6. B. Bollob´as。随机图中给定度数的顶点。《图论杂志》，6(2)：147–155，1982。7. B. Bollob´as。随机图。剑桥大学出版社，2001。8. S. Chatterjee。三角形计数大偏差中缺失的对数项。《随机结构与算法》，40(4)：437–451，2012。9. S. Chatterjee。随机图大偏差简介。《美国数学会通报》，53(4)：617–642，2016。10. S. Chatterjee 和 A. Dembo。非线性大偏差。《数学进展》，299：396–450，2016。11. S. Chatterjee 和 S. Varadhan。Erd˝os-R´enyi 随机图的大偏差原理。《欧洲组合学杂志》，32：1000–1017，2011。12. N. Cook 和 A. Dembo。稀疏 Erd˝os-R´enyi 图的子图计数大偏差。《数学进展》，373：107289，2020。13. B. Demarco 和 J. Kahn。三角形的上尾。《随机结构与算法》，40(4)：452–459，2012。14. H. D¨oring 和 P. Eichelsbacher。随机图中的中等偏差与伯努利随机矩阵的谱。《电子概率杂志》，14：2636–2656，2009。15. H. D¨oring 和 P. Eichelsbacher。通过累积量研究中等偏差。《理论概率杂志》，26：360–385，2013。16. R. Eldan。高斯宽度梯度复杂度、反向对数Sobolev不等式与非线性大偏差。《几何与泛函分析》，28：1548–1596，2018。17. V. F´eray、P. M´eliot 和 A. Nikeghbali。Mod-ϕ 收敛：正态性区域与精确偏差。施普林格概率与统计丛书，2016。18. D. Freedman。关于鞅的尾概率。《概率年刊》，3：100–118，1975。19. S. Ganguly、E. Hiesmayr 和 K. Nam。常数平均度随机图中三角形数量的上尾行为。《组合数学》，44(4)：699–740，2024。20. C. Goldschmidt、S. Griffiths 和 A. Scott。Erd˝os-R´enyi 随机图 $G(n, m)$ 和 $G(n, p)$ 中子图计数的中等偏差。《美国数学会会刊》，373：5517–5585，2020。21. M. Harel、F. Mousset 和 W. Samotij。通过高阶矩和熵稳定性研究上尾。《杜克数学杂志》，171(10)：2089–2192，2022。22. M. Harel、F. Mousset 和 W. Samotij。重新研究等差数列的上尾。arXiv:2409.*08383 [*math.PR*]*, 2024。23. W. Hoeffding。有界随机变量和的概率不等式。《美国统计学会杂志》，58(301)：13–30，1963。24. S. Janson。随机图的泛函极限定理及其在子图计数统计量中的应用。《随机结构与算法》，1(1)：15–37，1990。25. S. Janson。随机图统计量的正交分解与泛函极限定理。《美国数学会回忆录》，111，1994。26. S. Janson、T. Luczak 和 A. Ruci´nski。随机图。Wiley-Interscience，2000。27. S. Janson 和 K. Nowicki。广义 U 统计量的渐近分布及其在随机图中的应用。《概率论及相关领域》，90：341–375，1991。

28. S. Janson、K. Oleszkiewivz 和 A. Ruci´nski。《随机图中子图计数的上尾》。《以色列数学杂志》，142：61–92，2004。 29. S. Janson 和 A. Ruci´nski。《臭名昭著的上尾》。《随机结构与算法》，20(3)：317–342，2001。 30. J. Kim 和 V. Vu。《分治鞅与随机图中的三角形数量》。《随机结构与算法》，24(2)：166–174，2004。 31. K. Krokowski、A. Reichenbachs 和 C. Th¨ale。《离散 Malliavin-Stein 方法：随机图与渗流的 Berry-Esseen 界》。《概率年刊》，45：1071–1109，2017。 32. A. Liebenau 和 N. Wormald。《按度序列对图进行渐近计数，以及随机图的度序列》。《欧洲数学会杂志》，26(1)：1–40，2024。 33. E. Lubetzky 和 Y. Zhao。《随机图大偏差的副本对称性》。《随机结构与算法》，47，2012。 34. E. Lubetzky 和 Y. Zhao。《稀疏随机图上尾的变分问题》。《随机结构与算法》，50(3)：420–436，2017。 35. C. McDiarmid。《有界差分法》。载于《组合学综述》，第148–188页。剑桥大学出版社，1989。 36. B. McKay 和 N. Wormald。《按度序列对高阶图进行渐近计数》。《欧洲组合学杂志》，6：565–580，1990。 37. Joe Neeman、Charles Radin 和 Lorenzo Sadun。《环计数中的中等偏差》。《随机结构与算法》，63(3)：779–820，2023。 38. G. Reinert 和 A. R¨ollin。《随机子图计数与 U统计量：通过可交换对和嵌入进行多元正态逼近》。《应用概率杂志》，47：378–393，2010。 39. A. R¨ollin。《Erd˝os–R´enyi 随机图中三角形数量正态逼近的 Kolmogorov 界》。《工程与信息科学中的概率》，36(3)：747–773，2017。 40. A. Ruci´nski。《随机图的小子图何时服从正态分布？》。《概率论及相关领域》，78：1–10，1988。 41. V. Vu。《关于随机图中小子图数量的大偏差结果》。《组合、概率与计算》，10(1)：79–94，2001。

Appendix A. Proof of Proposition 2.7

在本节中，$\|x\|$ 表示欧几里得范数。

定义 A.1. 如果函数 $F\colon \mathbb{R}^d \to \mathbb{R}$ 是径向函数，则 $F(x)$ 的值仅取决于 $\|x\|$。如果 F 是径向函数，令 $F_{rad}\colon \mathbb{R}_0^+ \to \mathbb{R}$ 为满足以下条件的函数：对于所有满足 $\|x\| = r$ 的 x，都有 $F_{rad}(r) = F(x)$。

给定子集 $\mathcal{L} \subseteq \mathbb{Z}^d$，如果所有 $x \in \mathcal{L}$ 的所有坐标x_i 均非零，则称 $\mathcal{L}$ 具有非零性质。

命题 A.2. 设 $r \in \mathbb{R}$，并设 $F\colon \mathbb{R}^d \to \mathbb{R}^+$ 为连续、可积的径向函数，且 F_{rad} 为非递增函数。此外，设 $\mathcal{L}$ 为具有非零性质的 $\mathbb{Z}^d$ 的子集。则，

$$\sum_{x\in\mathcal{L}:\|x\|\geqslant r} F(x) \leqslant \int_{A(r-d^{1/2})} F(u)du$$

其中 $A(r) := \mathbb{R}^d \setminus B(0, r)$。

证明。我们不妨假设对所有 $x \in \mathcal{L}$，$\|x\| \geqslant r$。现在，对于每个 $x \in \mathcal{L}$，令 x_- 为将 x 的每个坐标的绝对值减 1 后得到的点（即正坐标减 1，负坐标加 1）。请注意，这一定义依赖于非零性质。

令 C_x 为由 x 和 x_-定义的开立方体，其含义是：若 y 的每个坐标都介于 x 和 x_- 中对应的坐标之间，则 $y \in C_x$。显然，不同的 $x \in \mathcal{L}$ 对应的立方体 C_x互不相交，且所有立方体的体积均为 1。因此，由于 x 是 C_x 中范数最大的点，且 F_{rad} 是非递增的，我们有

$$\sum_{x\in\mathcal{L}} F(x) \;=\; \sum_{x\in\mathcal{L}} \int F(u)\,\mathbb{I}_{C_x}(u)du \;=\; \int F(u)\,\mathbb{I}_{\bigcup_{x\in\mathcal{L}} C_x}(u)du\,.$$

由于单位立方体的直径为 $d^{1/2}$，且对所有 $x \in \mathcal{L}$都有 $\|x\| \geqslant r$，并集中的每个点的范数都至少为 $r - d^{1/2}$。由单调性即可得到所需结论。□

如果一个族 $\mathcal{L} \subseteq \mathbb{Z}^d$ 不具有非零性质，我们能说些什么呢？在这种情况下，可以根据非零坐标的集合 $S \subseteq [d]$ 对该族进行划分，并应用命题 A.2。由此得到以下推论。

推论 A.3. 设 $r \in \mathbb{R}$，且 $F\colon \mathbb{R}^d \to \mathbb{R}^+$ 是连续、可积的径向函数，并且 F_{rad} 非递增。那么

$$\sum_{x\in\mathbb{Z}^d:\|x\|\geqslant r} F(x) \;\leqslant\; 2^d \max_{S\subseteq[d]} \int_{A_S(r-d^{1/2})} F_S(u)du$$

其中 $A_S(r) := \mathbb{R}^S \setminus B(0,r)$，且 F_S 是 F 在 $\mathbb{R}^S$ 上的限制。

我们将对函数 $F(x) = \exp(-\beta\|x\|^2)$（其中 $\beta \geqslant 1$）应用这一结果。注意，给 $\int_{A(r)}\exp(-\beta\|u\|^2)du$ 找一个界并不困难。事实上，我们将利用：对于满足 $\|u\| \geqslant r$ 的 $u \in \mathbb{R}^d$，有 $\|u\|^2 \geqslant r^2/2 + (u_1^2 + \ldots + u_d^2)/2$。因此

$$\begin{aligned}
\int_{A(r)} \exp(-\beta\|u\|^2) &\leqslant e^{-\beta r^2/2} \int_{A(r)} \exp(-\beta(u_1^2+\cdots+u_d^2)/2) \\
&\leqslant e^{-\beta r^2/2} \int_{\mathbb{R}^d} \exp(-(u_1^2+\cdots+u_d^2)/2)du \\
&\leqslant e^{-\beta r^2/2} \prod_{i=1}^{d} \int \exp(-u_i^2)du_i \\
&= (2\pi)^{d/2} e^{-\beta r^2/2}\,.
\end{aligned}$$

请注意，在最后两行中，我们使用了富比尼定理和著名恒等式 $\int e^{-y^2/2}dy = \sqrt{2\pi}$。

显然，在一个 s 维子空间中进行相同的计算，将得到上界 $(2\pi)^{s/2}e^{-\beta r^2/2}$。所需的界，

$$\sum_{x\in\mathbb{Z}^d:\|x\|\geqslant r} \exp\left(-\beta\|x\|^2\right) \leqslant (8\pi)^{d/2}e^{-\beta(r-d^{1/2})^2/2}$$

现在由这些估计和推论 A.3 得出。命题 2.7 的证明至此完成。

Appendix B. The degree sequence – proofs from Section 3

我们将证明第3节中陈述的引理 3.1 和命题 3.2。先从引理 3.1 开始，我们在此重述如下。

引理 3.1. 设 $t \geqslant 2n^{-1}\log n$ 且令 $b \geqslant 4tn$。那么，除概率至多为 $\exp(-b)$ 的情形外，我们有

(i) $\Delta(G_m) \leqslant 2b/\ell_b$, *and*

(ii) $|V_j| \leqslant \dfrac{b}{tnj2^{j-6}}$ *for all* $j \geqslant 5$.

证明. 固定 $t \geqslant 2n^{-1}\log n$ 和 $b \geqslant 4tn$。注意，$b \geqslant 8\log n$。(i) 设 $a = 2b/\ell_b$。共有 n 个顶点，且对每个顶点，概率至多为

$$\binom{n}{a}t^a \leqslant \exp\left(-a\log(a/etn)\right) \leqslant \exp\left(\frac{-2b}{\ell_b}(\ell_b - \log\ell_b)\right) \leqslant \exp(-3b/2)$$

其度至少为 a。由于对每个顶点，这至多为 $\exp(-b)/2n$，因此条目 (i) 以至多 $\exp(-b)/2$ 的概率失效。

(ii) 我们将证明，对每个 $j \geqslant 5$，都有 $\mathbb{P}(|V_j| \geqslant b/tnj2^{j-6}) 6 \exp(-2b)$。由并集界，证明即告完成。固定 $j \geqslant 4$。必要时增大 b （这对应于一个略强的界），我们可以假设 $a := b/tnj2^{j-6}$ 是整数。

若事件 $|V_j| \geqslant a$ 发生，则存在一个由 a 个顶点组成的集合 A，其度数之和至少为 $2^j atn$。令 $G[A]$ 表示由集合 A 诱导的图，并令 $G[A,B]$ 表示集合 A 和 B 之间的二部诱导子图。由此可知，以下情况至少有一种成立：

(i) $e(G_m[A]) \geqslant 2^{j-2}atn$, or
(ii) $e(G_m[A, V\setminus A]) \geqslant 2^{j-1}atn$

在两种情况下，这个随机变量 X 都服从超几何分布，因此我们可以使用引理 2.1 中给出的界 (2.4)。注意，在这两种情况下都有 $\mu \leqslant atn$，因此令 $\nu = atn$，我们得到

$$\mathbb{P}\left(X \geqslant 2^{j-2}atn\right) \leqslant \exp(-(j-2)2^{j-4}atn) \leqslant \exp(-2b)\,,$$

如所要求。 □

现在我们转向命题 3.2，并将其重新表述。我们回顾一下定义。

$$\kappa(b,t) := \begin{cases} tn^2 & 1 \leqslant b < t^{1/2}n\ell \\ b^2/\ell^2 & t^{1/2}n\ell \leqslant b < n\ell \\ bn/\ell & n\ell \leqslant b \leqslant tn^2\ell \end{cases}$$

and

$$\kappa^+(b,t) := \begin{cases} b^2/\ell_b^2 & 1 \leqslant b < n\ell \\ bn/\ell & n\ell \leqslant b \leqslant tn^2\ell \end{cases}$$

命题 3.2. 存在一个绝对常数 $C > 0$，使得以下结论成立。假设 $t \geqslant 2n^{-1}\log n$ 且 $b \geqslant 32tn$。除去概率至多为 $\exp(-b)$ 的情况，我们有

$$\sum_u D_u(G_i)^2 \leqslant C\kappa(b,t)$$

对所有步骤 $i \leqslant m$，以及

$$\sum_{u\,:\,d_u(G_m)\geqslant 32tn} d_u(G_m)^2 \leqslant C\kappa^+(b,t)\,.$$

证明。固定 $t \geqslant 2n^{-1}\log n$，以及 $i \leqslant m$。我们证明这些界失效的概率至多为 $\exp(-2b)$。由并集界即可得到该命题。我们先考虑第二个结论，即关于受限和 $\sum_{u:\ d_u(G_m)\geqslant 32tn} d_u(G_m)^2$ 的结论。我们将证明，该结论可由引理 3.1 中关于 V_j 的界推出。定义 $J = J(b)$ 为满足以下条件的最大值：若 $b \leqslant n\ell$，则 $2^j tn \leqslant 2b/\ell_b$；若 $b > n\ell$，则 $J = \log_2(1/t)$。由引理 3.1，除去概率至多为 $\exp(-2b)$ 的情况，我们有

(i) 对所有 $j \geqslant 5$，$|V_j| \leqslant b/tnj2^{j-7}$；
(ii) 对所有 $j > J$，$V_j = \emptyset$。

对于范围 $32tn \leqslant b \leqslant n\ell$：假设 (i) 和 (ii)，我们有

$$\begin{aligned} \sum_{u:d_u(G_m)\geqslant 32tn} d_u^2(G_m) &\leqslant \sum_{j=5}^{J} 2^{2j+2}t^2n^2|V_j| \\ &\leqslant 2^9 \sum_{j=5}^{J} \frac{2^j btn}{j} \\ &\leqslant \frac{2^{J+11}btn}{J} \\ &\leqslant \frac{2^{12}b^2}{\ell_b^2}\,. \end{aligned}$$

其中在最后一个不等式中我们用到了 $2^j \geqslant b/tn\ell_b$，因此尤其有，

$$J \geqslant \log\left(\frac{b}{etn\ell_b}\right) = \ell_b - \log(\ell_b) \geqslant \frac{\ell_b}{2}.$$

这就完成了在此范围内关于 $\sum_{u:\ d_u(G_m)\geqslant 32tn} d_u(G_m)^2$ 的界的证明，前提是 C 至少取为 2^{12}。

对于范围 $b \geqslant n\ell$：在这种情况下， $J = \log_2(1/t)$，因此 $\lfloor J\rfloor \geqslant \ell$ 且 $2^{\lfloor J\rfloor} \leqslant t^{-1}$。假设 (i) 和 (ii) 成立，我们可以像上面那样推理，得到

$$\begin{aligned}\sum_{u:d_u(G_m)\geqslant 32tn} d_u(G_m)^2 &\leqslant \sum_{j=5}^{\lfloor J\rfloor} 2^{2j+2}t^2n^2|V_j| \\ &\leqslant \frac{2^{\lfloor J\rfloor+11}btn}{\lfloor J\rfloor} \\ &\leqslant \frac{2^{11}bn}{\ell}\,.\end{aligned}$$

只要 $C \geqslant 2^{11}$，这就证明了该范围内所需的界。

现在我们证明无约束和 $\sum_u D_u(G_i)^2$ 的第一个界。注意，在第一个范围内，$\kappa(b, t)$ 不依赖于 b，因此该范围内的结论可由下一个范围的开头推出（因为增大 b 对应于更强的结论）。此外，我们可以继续假设证明第一部分中已证明的各个界。特别地，若 $d_u(G_i) \geqslant 32tn$，则 $d_u(G_m) \geqslant 32tn$，并且有 $D_u(G_i)^2 \leqslant d_u(G_m)^2$。

在范围 $t^{1/2}n\ell \leqslant b < n\ell$ 内，由上述界可得

$$\sum_{u:d_u(G_i)\geqslant 32tn} D_u(G_i)^2 \leqslant \frac{2^{12}b^2}{\ell_b^2}$$

根据引理 3.3，存在常数 C_1，使得除概率为 $\exp(-2b)$ 的情形外，我们有 $\sum_{u:d_u(G_i)\leqslant 32tn} D_u(G_i)^2$ 在 $b \leqslant n$ 时至多为 Ctn^2，而在 $n \leqslant b \leqslant n\ell$ 时至多为 $Cbtn$。由于 $\max\{tn^2, btn, b^2/\ell^2\} = b^2/\ell^2 = \kappa(b,t)$，我们有

$$\sum_u D_u(G_i)^2 \leqslant (2^{12} + C_1)\kappa(b,t)\,,$$

这就得到了该范围内的结果。

在范围 $n\ell \leqslant b < tn^2\ell$ 内，我们同样将两项分别界定为 $2^{11}bn/\ell$ 和 C_1btn。若选择 $C \geqslant 2^{11} + C_1$，则其至多为 $C\kappa(b,t)$。□

Appendix C. Variance lower bound

我们将证明引理 7.3，并为方便起见重新陈述该引理。我们还回顾定义

$$\mathbb{X}_i'' := 3\frac{(N-m)_2(m-i)}{(N-i)_3}\,X_\wedge''(G_i) + \frac{(N-m)_3}{(N-i)_3}\,X_\triangle''(G_i)\,,$$

引理 7.3。存在一个绝对常数 $c_1 > 0$，使得以下结论成立。设 $n^{-1/2}/c_1 \leqslant t \leqslant 1/2$。对于所有 $m/2 \leqslant i \leqslant m$，我们有

$$\mathbb{P}\left(\mathbb{E}\left[(\mathbb{X}_i'')^2 | G_{i-1}\right] \leqslant c_1 t^2 n\right) \leqslant \exp(-c_1 n)$$

对所有充分大的 n。

我们指出，若 t 是常数，即在稠密随机图的情形下，则可直接由 [20] 第7节给出的方差界推出。由于 $X''_{\wedge}(G_i)$ 的系数为 $O(t)$，而 $X''_{\triangle}(G_i)$ 的系数为 $\Omega(1)$，显然只需证明，除概率为 $\exp(-\Omega(n))$ 的情况外，我们有

$$\mathbb{E}\left[X''_{\wedge}(G_i)^2 \mid G_{i-1}\right] = O\left(\frac{n}{\ell^2}\right) \tag{C.1}$$

and

$$\mathbb{E}\left[X''_{\triangle}(G_i)^2 \mid G_{i-1}\right] = \Omega(t^2 n)\,. \tag{C.2}$$

注意，(C.1) 可由引理 3.4 推出（因为截断只会减小方差），因此我们只需证明，除概率为 $\exp(-\Omega(n))$ 的情况外，(C.2) 成立。

$X_{\triangle}(G_i)$ 的下一项中的方差来自图中不同的共度（若所有非边 uw 的共度 $d_{uw}(G_{i-1})$ 都相等，则方差为零）。回顾记号 $D_{uw}(G_m)$，它表示点对 uw 相对于其在 $G_m \sim G(n,m)$ 中期望值的共度偏差。由 (5.2) 可得

$$X_{\triangle}(G_i) = 6D_{e_i}(G_{i-1}) - \mathbb{E}\left[6D_{e_i}(G_{i-1}) \mid G_{i-1}\right] \tag{C.3}$$

其中 e_i 表示过程的第 i 条边。为了将其转化为对共度的比较，我们回顾一个基本事实：对于均值为 0 的独立同分布随机变量 X、Y，有 $\mathbb{E}\left[(X-Y)^2\right] = 2\mathbb{E}\left[X^2\right]$。类似地，若 X、Y 是满足 $\mathbb{E}\left[X|G_{i-1}\right] = 0$ 的独立同分布随机变量，则 $\mathbb{E}\left[(X-Y)^2|G_{i-1}\right] = 2\mathbb{E}\left[X^2|G_{i-1}\right]$。在 $X_{\triangle}(G_i)$ 的情况下，随机性来自边 $e_i \in E(K_n) \setminus E(G_{i-1})$ 的选取。因此可得

$$\begin{aligned}
\mathbb{E}\left[X_{\triangle}(G_i)^2 \mid G_{i-1}\right] &= \frac{1}{2(N-i+1)^2} \sum_{e',e^* \notin E(G_{i-1})} \left(X_{\triangle}(G_{i-1}\cup\{e'\}) - X_{\triangle}(G_{i-1}\cup\{e^*\})\right)^2 \\
&= \frac{18}{(N-i+1)^2} \sum_{e',e^* \notin E(G_{i-1})} \left(D_{e'}(G_{i-1}) - D_{e^*}(G_{i-1})\right)^2 \\
&= \Omega(n^{-4}) \sum_{e',e^* \notin E(G_{i-1})} \left(D_{e'}(G_{i-1}) - D_{e^*}(G_{i-1})\right)^2 .
\end{aligned}$$

然而，这尚未考虑截断；不过，截断只会影响大于 $3t^2 n$ 的偏差，因此，通过同样的论证，

$$\mathbb{E}\left[X''_{\triangle}(G_i)^2 \mid G_{i-1}\right] = \Omega(n^{-4}) \sum_{e',e^* \in E_0(G_{i-1})} \left(D_{e'}(G_{i-1}) - D_{e^*}(G_{i-1})\right)^2$$

其中 $E_0(G) = \{uw \notin E(G_{i-1}):\ D_{uw}(G_{i-1}) \leqslant 3t^2 n\}$。

因此，只需证明：除概率为 exp(−Ω(n)) 的情况外，至少存在 Ω(n^2) 个点对 $uw \in E(K_n) \setminus E(G_{i-1})$，且满足以下条件。

$$d_{uw}(G_{i-1}) \leqslant sd_{med}(G) - sn^{1/2} \tag{C.4}$$

以及 Ω(n^2) 个点对 $uw \in E(K_n) \setminus E(G_{i-1})$，满足

$$sd_{med}(G) + sn^{1/2} \leqslant d_{uw}(G_{i-1}) \leqslant 1.1s^2n + 2sn^{1/2} \tag{C.5}$$

其中 $d_{med}(G)$ 是中位度。注意，这给出了 Ω(n^4) 个点对的点对 $uw, u'w' \in E_0(G_{i-1})$，满足 $|D_{e'}(G_{i-1}) - D_{e*}(G_{i-1})| \geqslant 2sn^{1/2}$，这显然足够（因为$t/2 \leqslant s \leqslant t$）。

现在我们将证明，对于 Ω(n^2) 个点对 $uw \in E(K_n) \setminus E(G_{i-1})$，除概率为 exp(−Ω($n$)) 的情况外，(C.5) 成立。显然，除了一些细微变化外，同一证明也能推出 (C.4)。

我们先给出一个引理，从单个顶点 u 的角度考察共度 $d_{uw}(G)$。我们考虑 $G \sim G(n,p)$，并观察到它可以推广到我们的情形。对于一个顶点 $u \in V(G)$，令 F_u 为以下事件：

(i) $d(u) \in [0.9pn, 1.1pn]$
(ii) $|\{w \in V(G) \setminus N(u) : d_{uw}(G) \in [pd(u) + pn^{1/2}, pd(u) + 2pn^{1/2}]\}| < n/100$

引理 C.1. 存在一个常数 $c_1 > 0$，使得以下结论成立。设 $n^{-1/2}/c_1 \leqslant t \leqslant 1/2$，设 $p \in (t/2, t)$，并设 $G \sim G(n,p)$。那么，对于每个顶点$u \in V(G)$，我们有

$$\mathbb{P}(F_u) \leqslant \exp(-c_1 n).$$

证明。如果事件 F_u 发生，那么我们必定有 $d(u) \in [0.9pn, 1.1pn]$。因此，如果我们考虑以 $N(u) = S$ 为条件化，其中 S 是某个顶点集合，那么我们有

$$\mathbb{P}(F_u) = \sum_{0.9pn \leqslant |S| \leqslant 1.1pn} \mathbb{P}(N(u) = S) \cdot$$

$$\mathbb{P}\left(|\{w \in V(G) \setminus S : |N(w) \cap S| \in [p|S| + pn^{1/2}, p|S| + 2pn^{1/2}]\}| < n/100\right),$$

其中我们已经利用了 $G(n,p)$ 中的边相互独立，因此 w 与 S 之间的边独立于 S 是 u 的邻域这一事实。由于概率之和 $\mathbb{P}(N(u) = S)$ 至多为 1，且所有基数相同的 S 都是等价的，因此只需证明，对于范围 $0.9pn \leqslant r \leqslant 1.1pn$ 内的所有整数 r，事件

$$\left|\{w \in V(G) \setminus \{1, \dots r\} : \left|N(w) \cap \{1, \dots, r\}\right| \in [pr + pn^{1/2}, pr + 2pn^{1/2}]\}\right| < n/100$$

其概率至多为 exp(−Ω(n))。对于每个固定的顶点 $w \in V(G)$，随机变量 $B_{w,\,r} := |N(w)$ 的分布为 $Bin(r,p)$。中心极限定理指出，$(B_{w,\,r} - rp)/(rp(1-p))^{1/2}$ 在 $r \to \infty$ 时收敛于标准正态分布（因此特别地，在 $n \to \infty$ 时也成立）。并且，令 $\sigma = (rp(1-p))^{1/2}$，我们得到

即 $pn^{1/2} = \lambda\sigma$，其中 $\lambda = (pn/r(1-p))^{1/2}$，并且可以验证 $\lambda \in [0.6, 1.1]$。由此可知，$B_{w,\ r} \in [pr + pn^{1/2}, pr + 2pn^{1/2}]$ 的概率收敛到 $\Phi(2\lambda) - \Phi(\lambda)$，其中某个 $\lambda \in [0.6, 1.1]$，且该值总是至少为 $1/10$。

由于当我们改变 $w \in V(G) \setminus \{1, \dots r\}$ 的取值时，该事件相互独立，因此我们得到 $|\{w \in V(G) \setminus \{1, \dots r\} : \ |N(w) \cap \{1, \dots, r\}| \in [pr + pn^{1/2}, pr + 2pn^{1/2}]\}|$ 随机占优一个服从分布 $Bin(n - r, 1/10)$ 的随机变量，因此也随机占优 $Y \sim Bin(n/3, 1/10)$。所需的界现在即可得出，因为 Chernoff 不等式给出 $\mathbb{P}(Y < n/100) \leqslant \exp(-\Omega(n))$。□

该引理显然对于证明 $\Omega(n^2)$ 点对$uw \in E(K_n) \setminus E(G_{i-1})$ 满足 (C.5) 非常有用。我们立即注意到，除了一个多项式因子外（它很容易被吸收到 $\exp(-\Omega(n))$ 项中），同样的结果也适用于 G_{i-1}，而且我们还可以对顶点 u 使用并集界。因此，除概率至多为 $\exp(-\Omega(n))$ 的事件外，我们有

$$|\{w \in V(G_{i-1}) \setminus N(u) : \ d_{uw}(G) \in [sd(u) + sn^{1/2}, sd(u) + 2sn^{1/2}]\}| < n/100 \qquad \text{(C.6)}$$

对于所有度处于区间 $[0.9sn, 1.1sn]$ 内的顶点 $u \in G_{i-1}$。

根据命题 3.2，图 G_i 中有线性数量的顶点的度不在范围 $[0.9sn, 1.1sn]$ 内的概率至多为 $\exp(-\Omega(s^2n^2\ell)) \leqslant \exp(-\Omega(n))$。因此，我们可以假设至少有 $n/10$ 个顶点 u 满足 $0.9sn \leqslant d_{med}(G) \leqslant d_u(G_{i-1}) \leqslant 1.1sn$，并且对于这样的顶点，至少有 $n/100$ 个顶点 w，由 (C.6) 给出，使得

$$sd_{med}(G) + sn^{1/2} \leqslant d_{uw}(G_{i-1}) \leqslant 1.1s^2n + 2sn^{1/2}.$$

I由此立即可知，(C.5) 对 $\Omega(n^2)$ 个点对 $uw \in E(K_n) \setminus E(G_{i-1})$ 成立。

如上所述，类似的论证表明，对于 $\Omega(n^2)$ 点对 $uw \in E(K_n) \setminus E(G_{i-1})$，(C.4) 也成立。至此，引理 7.3 的证明完成。

Faculty of Mathematics and Physics, University of Ljubljana, Jadranska ulica19, 1000 Ljubljana, Slovenia
电子邮箱地址：jose.alvarado@fmf.uni-lj.si

Departamento de Matemática, PUC-Rio, Rua Marquês de São Vicente 225, Gávea,22451-900 Rio de Janeiro, Brazil

Departamento de Matemática, PUC-Rio, Rua Marquês de São Vicente 225, Gávea,22451-900 Rio de Janeiro, Brazil
电子邮箱地址：simon@puc-rio.br

本研究部分由巴西高等教育人员提升协调会（CAPES）提供资金支持，财务代码001。