arXiv:2305.02974v2 [math.CO] 7 Jul 2023

关于具有有限依赖关系的随机图模型中的连通性

约翰内斯·伦格勒[1],安德斯·马丁松[1],卡莉娜·彼得罗娃[1],帕特里克·施奈德[1],拉斐尔·施泰纳[1],西蒙·韦伯[1],和埃莫·韦尔茨尔[1]

[1]计算机科学系，苏黎世联邦理工学院¨

摘要

对于任意正边密度 p，Erd˝os-R´enyi $G_{n,p}$ 模型中的随机图以非零概率连通，因为所有边都相互独立。我们考虑这样一类随机图模型：不共享端点的边相互独立，而关联边之间可能存在依赖关系。我们提出以下问题：最小概率 $\rho(n)$是多少，使得对于该模型中任意一个定义在具有 n 个顶点的图上的分布 $\mathcal{G}$，只要每条候选边存在的边缘概率至少为 $\rho(n)$，从 $\mathcal{G}$ 中抽取的图就以非零概率连通?

事实证明，“不共享端点的边彼此独立”这一条件需要进一步明确，而上述问题的答案取决于具体的规定。事实上，我们将这一直观描述形式化为由五种独立性条件构成的严格层级，并证明就阈值 $\rho(n)$ 而言，它们至少呈现出三种不同的行为。对于每种条件，我们都给出 $\rho(n)$ 的上界和下界。在最强的条件下，即着色模型（例如包括随机几何图），我们证明 $\rho(n) \to 2-\phi \approx 0.38$ 对于 $n \to \infty$，从而证明了 Badakhshian、Falgas-Ravry 和 Sharifzadeh 提出的猜想。这使着色模型区别于我们所考察的较弱独立性条件；对于后者，我们证明 $\rho(n) > 0.5 - o(n)$。与着色模型形成鲜明对比的是，对于我们最弱的独立性条件——不相邻的边两两独立——我们证明，对于完全任意的分布， $\rho(n)$ 与阈值$1 - 2/n$ 相差不超过 $O(1/n^2)$。

致谢。 本研究始于2023年1月在瑞士施泰尔斯举行的苏黎世联邦理工学院（ETH Zürich）组合结构与算法研究组和组合算法理论研究组的联合研讨会。我们感谢组织者为我们提供了十分愉快且富有启发性的研究氛围。K.P. 获得了瑞士国家科学基金会第 CRSII5 173721 号资助。R.S. 获得了苏黎世联邦理工学院博士后奖学金。S.W. 获得了瑞士国家科学基金会项目第 204320 号的资助。最后但同样重要的是，我们要感谢 Victor Falgas-Ravry 指出了大量相关研究成果，从而极大地帮助我们更好地理解本研究成果的定位。

1 引言

概率方法是理论计算机科学、图论和组合学中的重要工具 [3]。利用这种方法，可以证明某个随机构造以正概率具有某种理想性质，从而得出必定存在具有该性质的某些对象。概率方法的早期例子包括构造扩展图 [8]，或具有较大围长和较大色数的图 [17]。通常，可以将该性质表示为若干非常简单的事件的交集 $A_1, \ldots, A_k$。例如，我们可以将一个集合 S 是团这一性质，表示为对所有点对$u, v \in S$而言，事件"u 与 v相邻"的交集。在许多情况下，显然每个 A_i 发生的概率都很大，但通常仅凭这一点，还不足以得出所有 A_i 同时发生的概率为正。当然，如果概率空间是一个乘积空间，其中所有分量彼此独立，那么这就显而易见了。

然而，对于许多应用而言，将自身局限于乘积空间的做法限制性过强。通过认识到概率方法仍然可以应用于不具备完全独立性的情形，只要系统中的依赖程度受到某种约束，概率方法便得到了显著拓展。其开创性成果是 Lov´asz 局部引理（LLL），我们在此给出其稍强的 Shearer [30] 版本：对于事件 $A_1, \ldots, A_k$，若它们发生的概率均至少为 p，且每个 A_i 至多依赖于其他 d 个 A_j，其中 $p \geq 1 - 1/(ed)$，那么所有 A_i 同时发生的概率为正。[1] 在这一背景下，依赖关系由一个依赖图来刻画，该图的顶点集为 $\{A_1, \ldots, A_k\}$，其中每个顶点都与除其邻居之外的所有顶点相互独立。在许多情况下，依赖图可以受到限制，我们将在下文第 1.1 节中给出一些例子。

局部引理允许通过后门重新引入乘积空间，因为局部引理条件 $p \geq 1 - 1/(ed)$ 允许将该过程与乘积空间耦合。这一点已隐含在局部引理的归纳证明 [31], 中，后来由 Liggett、Schonmann 和 Spacey [25]明确阐述。他们还将这种耦合推广到可数无穷多个变量，并证明了 Shearer 条件的紧性（脚注 1 中精确版本的紧性）。对于有限个变量，将概率空间与乘积空间耦合，便可直接推出全为一事件具有正概率。

遗憾的是，局部引理在 d方面的尺度表现不佳；也就是说，若 d 很大，p 就必须非常接近 1。然而，在某些情况下，度数 d 可能会增长。具体而言，本文将研究这样一种情形：变量与一个具有 n 个顶点的完全图的边相关联，且依赖关系仅存在于相邻的边之间（即共享一个端点的边之间）。这是一种常见情形，例子见第 1.1 节。在这种情况下，每条边的依赖数量为 $2(n-2)$，因此会随着 n增长。因此，如果每条边以固定概率 $p < 1$出现，那么当 n较大时，就无法将概率空间与乘积空间耦合，也无法保证完全图以正概率出现 [25]。不过，本文将证明，可以保证某些较弱的全局性质 。具体而言，我们将关注连通性，因为这可以说是最基本的全局图性质。我们将研究以下问题：

> 考虑一个随机图，其中每条候选边都以至少 p 的概率加入。假设不相邻的边相互独立。对于哪些 p 的取值，我们可以保证该图以正概率连通?

结果表明，答案在很大程度上取决于我们对独立性的具体定义。令

[1]精确条件为 $p > 1 - \frac{(d-1)^{(d-1)}}{d^d}$ 对于 $d > 1$，以及 $p > 1/2$ 对于 $d = 1$。根据估计式 $\frac{(d-1)^{(d-1)}}{d^{d-1}} \geq e^{-1}$，第一个表达式可以进一步简化；当极限 $d \to \infty$ 时，该估计式变为紧确估计。

人惊讶的是，对于看似无害的条件“不相邻的边相互独立”，至少有五种不同的解读方式。我们接下来将按强度递增的顺序定义它们。对于一条边 $e = \{u, v\}$，令 X_e 表示 e 存在于随机图中的事件。

定义 1（两两独立性）。对于每一对不相邻的边 e, f，随机变量 X_e 和 X_f 相互独立。

定义 2（匹配独立性）。对于任意一组 M 两两不相邻的边（也称为匹配），$\{X_e \mid e \in M\}$ 相互独立。

定义 3（边子图独立性）。对于任意一条边 $e = \{u, v\}$、任意顶点集 W，其中u, $v \notin W$，以及在 W上任取的图 H_W，对于从该分布中采样得到的随机图 G，有

$$Pr[G[W] = H_W \text{ and } X_e] = Pr[G[W] = H_W] \cdot Pr[X_e].$$

定义 4（子图独立性）。对于任意两个不相交的顶点子集 V, W，以及分别在 V和 W上任取的图 H_V和 H_W，对于从该分布中采样得到的随机图 G，有

$$Pr[G[V] = H_V \text{ and } G[W] = H_W] = Pr[G[V] = H_V] \cdot Pr[G[W] = H_W].$$

定义 5（着色模型）。图分布由一组概率空间（Ω_v, Pr_v）和一组确定性函数 $f_{u,\,v}\colon \Omega_u \times \Omega_v \to \{0,1\}$ 构成；每个顶点 $v \in V$对应一个概率空间，每条边 $\{u, v\} \in \binom{V}{2}$对应一个计算 $X_{\{u,v\}}$的函数。如果每个概率空间（Ω_v, Pr_v）都是有限的，我们称 $\max_v |\Omega_v|$ 为该着色模型的颜色数。

我们指出，局部引理的标准证明要求边子图独立性。在这篇论文中，作者明确说明了如何将其耦合到乘积空间，并将子图独立性描述为所需的性质；但仔细检查证明可知，他们实际只用到了较弱的边子图独立性条件。

请注意，我们的一些模型此前曾以不同名称进行研究。我们在图 1 中总结了本文发表前或同期得到的连通性阈值界，并在第 1.1 节中更详细地讨论相关工作。我们关于较大 n 的主要结果总结在表 1 和图 2 中。我们在定理 7 中对所有 n 的结果作了更精细的阐述。并非所有情形下我们都得到了匹配的上界和下界，但如图 2 所示，我们的界表明，在上文讨论的五种独立性条件中，至少存在三种不同的阈值；若还包括

独立性条件	下界	上界
着色模型（两种颜色）	$1/4$	$1/4$
着色模型（一般情形）	0.381966 ...	0.381966 ...
子图独立性	$1/2$ [15, 定理 16]	$1/2$ [15, 定理。 16]
边子图独立性	$1/2$	$3/4$
匹配独立性	$1/2$	1
两两独立性	1	1

表 1：阈值的下界和上界：ρ 使得每个常数边概率 $> \rho$在所有充分大的 n中都以正概率保证连通性，而常数边概率 $< \rho$ 对于无穷多个 n不以正概率保证连通性。

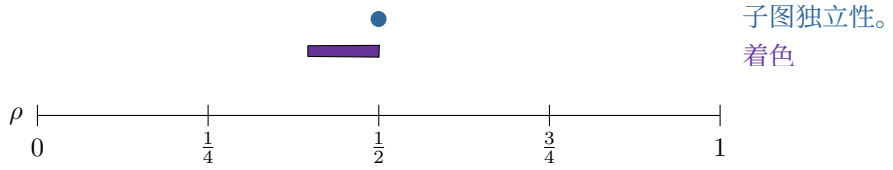


图 1：相关工作中已知阈值上下界的示意图。子图独立性的上下界见 [15, 定理 16]。着色模型的上下界是与我们的工作同时且独立地得到的，见 [5, 定理 1.7]。

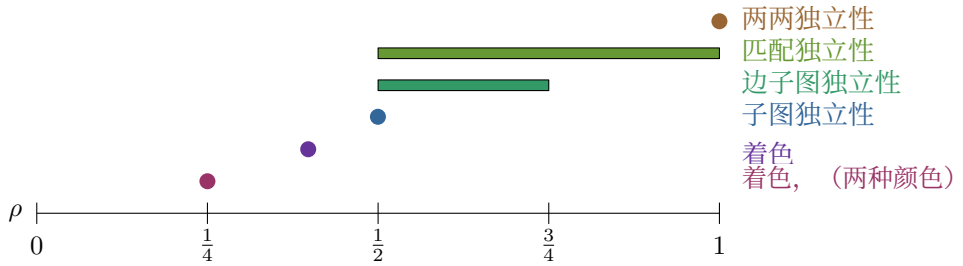


图 2：表 1 和定理 7 中所述上下界的示意图。可以看出，在六种独立性条件中，至少有四个阈值 ρ 。

仅有两种颜色的着色模型，则存在四种不同的阈值。值得注意的是，我们的证明为全部六种独立性条件的阈值给出了不同的区间，尽管这并不意味着这些阈值各不相同。

此外，记 $\mathcal{G}_{\mathrm{pw}}(n,\ p)$ 为所有定义在 n 个顶点上、边缘边概率至少为 p 且满足“两两独立性”的随机图分布所构成的类。记 $\mathcal{G}_{\mathrm{pw}} := \bigcup_{n\in\mathbb{N}} \mathcal{G}_{\mathrm{pw}}(n,\ 0)$ 为不限制顶点数量和边缘概率的同一类。类似地，我们用 $\mathcal{G}_{\mathrm{mat}}$、$\mathcal{G}_{\mathrm{esub}}$、$\mathcal{G}_{\mathrm{sub}}$ 和 $\mathcal{G}_{\mathrm{col}}$ 分别表示满足匹配独立性、边子图独立性、子图独立性以及着色模型的随机图分布。我们证明这五种模型构成一个严格的层级。

定理 6。 $\mathcal{G}_{col} \subsetneq \mathcal{G}_{sub} \subsetneq \mathcal{G}_{esub} \subsetneq \mathcal{G}_{mat} \subsetneq \mathcal{G}_{pw}$。

1.1 相关工作与示例

图论、理论计算机科学和组合学中有许多应用，其中依赖图的度数有界或受到其他限制。例如，如果我们想在超图中找到一种顶点着色，使其不存在单色超边，并且独立地为顶点着色，那么任意两个不相交的超边都是独立的。因此，对于最大度有界的图（超图）着色，通常可以应用局部引理类型的论证。例如，Johansson [22] 和 Molloy [26]提出的开创性结果指出，每个最大度为 Δ 的图的（列表）色数为 $O(\Delta/\log\Delta)$；该结果正是利用这种方法证明的。同样，如果我们想为合取范式的 SAT 公式找到一个满足赋值，并且独立地为变量赋值，那么只要两个子句不共享同一个文字，它们就会被独立地满足；更多细节和背景可参见 [19] 。这些都是局部引理的经典应用，属于着色模型这一类。

另一个重要的应用来源是渗流理论，这也是 [25]最初的主要动机。在该领域，研究局部生成的几何图的一项重要技术是重标度：几何空间被划分为可能部分重叠的盒子；若满足某些条件，则称盒子为良好，而这些条件仅取决于该盒子所诱导的子图。它

可以递归地证明，盒子成为良好盒子的概率会随着盒子尺寸的增大而提高。若两个盒子不重叠，则它们是否良好相互独立；通常每个盒子只与有限多个其他盒子重叠。

到目前为止，我们给出了有界依赖的一般示例。现在我们转向一些示例，其中随机变量自然地与完全图的边集联系在一起。一类重要的示例来自比 Erd˝os-R´enyi 图更复杂的随机图模型，尤其是那些能够捕捉现实世界网络的图性质的模型，例如聚类、社群或重尾度分布。许多此类模型通过为每个顶点抽取一些信息，并根据这些信息连接两个顶点来生成。此类随机图模型数量众多：在 Chung-Lu 或 Norros-Reittu 随机图中，顶点抽取决定其期望度的权重 [13, 29]；在随机几何图（RGG）或双曲随机图（HRG）中，顶点在几何空间中抽取某个位置 [24, 28]；在无标度渗流（SFP）中，顶点位于网格上，并且还抽取随机权重 [16]；在几何非齐次随机图（GIRG）中，它们同时抽取位置和权重 [10]。在随机块模型（SBM）中，它们抽取自己应当所属的社群 [23]。文献中还可以找到许多具有类似特点、无法在此一一列出的其他应用。所有这些模型都属于着色模型。[2]

着色模型这一类是最重要的一类。[3] 不属于这一类的模型可能出现在以下情形：从分布中抽取的图必须满足某些全局性质。例如，考虑如下针对顶点数为奇数的图的分布 $n \geq 5$。每个顶点都独立且均匀地随机选择红色或蓝色。所得图仅由所选颜色出现次数为偶数的顶点构成一个团。我们稍后会证明（在引理 16 的证明中），该分布具有匹配独立性，但不满足边子图独立性（因此也不是着色模型）。作为第二个例子，考虑 Erd˝os-R´enyi 模型 $G_{n,\ 1/2}$ 并附加如下约束：边的总数 $|E|$ 可被三整除。[4] 可以相当容易地证明，该分布甚至不满足两两独立性。另一方面，如果我们改用边的总数 $|E|$ 可被二整除这一附加约束，那么该分布实际上是一个着色模型，这表明独立性条件出人意料地反复无常。这两个事实的证明都可以在附录 A 中找到。

我们的两个独立性条件此前曾以不同名称被研究。首先，子图独立分布曾以 1-独立随机图之名进行研究。更一般地，a k-独立图分布是这样一种分布：如果两组边 $E,\ F$ 中任一条边的关联顶点与另一组中任一条边的关联顶点之间的最小距离至少为 k，那么这两组边相互独立。这些 k-独立图分布在渗流理论领域已得到广泛研究，例如在无限整数格 $[6]$ 上。在 $[15]$, Day、Falgas-Ravry 和 Hancock 的研究中，他们考察了 1-独立随机有限 图。在其他结果中，他们精确确定了 1-独立（即子图独立）随机图的连通性阈值，正如上文表 1 和图 1 所述。

着色模型一直以两种不同的名称进行研究。首先，在 $[15]$ 中的渗流研究背景下，它们被称为 基于顶点的测度 。其次，从确定性视角来看，

[2]许多模型的表述都涉及为每条边抛一次硬币，其纳入概率是两个端点信息的函数。在这种表述中，事件 X_e 并非信息的函数。不过，我们只需将抛硬币这一随机实验定义为其中一个端点的随机实验的一部分即可。[3]但请注意，这并不是自然导出局部引理的模型。这就引出了一个问题：能否针对着色模型获得更强的局部引理类型结果？[4]更形式化地说，我们考虑在事件 $|E| \equiv_3 0$ 条件下，对 $G_{n,\ 1/2}$ 分布进行条件化所得到的条件概率分布。

着色模型方面的研究核心在于广为研究的多部图的密度 *Tur´án* 问题 [7, 9, 14, 18, 27]: 多部图的密度 Tur´án 问题不考虑随机图分布，而是考虑图 G 的 $|V(G)|$-部膨胀图 G^*，以及 G^* 各部之间为保证图 H 作为横截子图而所需的最小密度，即作为 G^* 的子图，且每个部恰好选取一个顶点 G^*。对于不只考虑单个图H 作为横截子图，而是考虑某个集合 $\mathcal{H}$中的任意图这一更一般的情形，也有人研究过该问题。此外，还可以考虑加权情形，其中 G^* 的每个部中的顶点都被赋予正权重，且每个部中的所有权重之和为 1。于是，两部之间的密度也以加权方式计算。现在，将 G 视为完全图 K_n，将 $\mathcal{H}$ 视为 n个顶点上的生成树族，可以看出，多部图的加权密度 Tur´án 问题等价于确定在着色模型中为以非零概率保证连通性所需的最小边概率这一问题（不同之处在于，多部图的加权密度 Tur´án 设置只允许对有限多个颜色进行建模；正如我们将在下一小节中讨论的，这并不失一般性）。在与本文同时且独立发表的 [5], 中，Badakhshian、Falgas-Ravry 和 Sharifzadeh 恰好考虑了这一问题；他们证明了与我们相同的下界 [5, Thm. 1.7], ，并将我们的上界作为猜想提出 [5, Conj. 1.9]。

最后但同样重要的是，让我们提及，在一个概念上相似（但具体情形相当不同）的方向上，Alon 和 Nussboim [2] 研究了随机图模型的连通性阈值：在这些模型中，只有大小至多为 k 的边集（对于固定参数 k ）需要相互独立，而大小从 $k+1$ 开始的边集之间则可能存在依赖关系。

1.2 详细结果

In 本节给出主要定理及更详细的结果。我们将 b y

$$\rho_{\mathrm{pw}}(n) := \inf\{p \in [0,1] \mid \forall \mathcal{D} \in \mathcal{G}_{\mathrm{pw}}(n,p), \Pr[G_n \sim \mathcal{D} \text{ is connected}] > 0\}$$

能够保证 G_n 连通的最小（下确界）边的边缘概率。量 $\rho_{\mathrm{mat}}(n)$、 $\rho_{\mathrm{esub}}(n)$、 $\rho_{\mathrm{sub}}(n)$ 和 $\rho_{\mathrm{col}}(n)$ 的定义类似。此外，令 $\mathcal{G}_{\mathrm{col},\ k}$ 表示至多有 k 种颜色的着色模型，并令 $\rho_{\mathrm{col},\ k}(n)$ 表示相应的阈值。

定理 7. 我们有关于 ρ 的以下界:

两两独立性：对于所有 $n \in \mathbb{N}$,

$$1 - 2/n - \Theta(1/n^2) \leq \rho_{pw}(n) \leq 1 - 2/n.$$

匹配独立性：对于所有 $n \in \mathbb{N}$,

$$\frac{1}{2}(1 - \tan^2 \frac{\pi}{2n}) \leq \rho_{mat}(n) \leq 1 - 2/n.$$

此外，对于任意 $k \in \mathbb{N}$,

$$\frac{1}{2} \leq \rho_{mat}(8k).$$

边子图独立性：对所有 $n \in \mathbb{N}$,

$$\frac{1}{2}(1 - \tan^2 \frac{\pi}{2n}) \leq \rho_{esub}(n) \leq 3/4.$$

子图独立性： *[15,* 定理 *16]* 对所有 $n \geq 2$,

$$\rho_{sub}(n) = \frac{1}{2}(1 - \tan^2 \frac{\pi}{2n}).$$

着色模型：对于所有 $n \geq 3$，我们有 $1/4 \leq \rho_{col}(n) \leq 1/2$。此外，令 $\phi = \frac{1}{2}(1+\sqrt{5})$ 为黄金比例。那么

$$\lim_{n\to\infty} \rho_{col}(n) = 2 - \phi \approx 0.381966.$$

着色模型，两种颜色：对于所有 $n \geq 3$， $\rho_{col,\ 2}(n) = 1/4$。

据我们所知，此前尚未针对两两独立性、匹配独立性或边子图独立性模型研究过这个问题。对于子图独立性模型，其阈值已被精确确定为 [15], ，但为了完整呈现我们模型层级中的结果，我们仍将其纳入定理。我们关于着色模型的结果证实了 Badakhshian、Falgas-Ravry 和 Sharifzadeh 的一个猜想 [5]。

我们指出，尽管 $\mathcal{G}_{\mathrm{col}}(n,p)$ 包含在每个顶点进行的随机实验有无限多种结果的模型，但已有研究表明[5] （对于 $n = 3$ ，见 [9] ；对于所有n，见引理 2.1] ），$\rho_{\mathrm{col},k}(n) \geq \rho_{\mathrm{col}}(n)$ 对于 $k = n - 1$ 成立。结合显然的 $\rho_{\mathrm{col},k}(n) \leq \rho_{\mathrm{col}}(n)$ 对任意 k 成立，可得 $\rho_{\mathrm{col},n-1}(n) = \rho_{\mathrm{col}}(n)$。证明 $\rho_{\mathrm{col},n-1}(n) \geq \rho_{\mathrm{col}}(n)$ 时，采用了一个连通概率为 0 的模型，并将其修改为至多使用 $n-1$ 种颜色，同时保持其永不连通这一性质，并使其边缘边概率增大或保持不变。然而，修改后的模型在 n 个顶点上的所有图上诱导的概率分布通常不同于原模型。这里，我们证明一个更强的结论：对于某个 $k = k(n)$， $\mathcal{G}_{\mathrm{col}}(n,p) = \mathcal{G}_{\mathrm{col},k}(n,p)$，这表明有限种颜色足以模拟任何能够表示为着色模型的所有图上的概率分布。引理 8 的证明推迟到附录 E。

引理 8. 考虑一个定义在 n 个顶点上的着色模型，以及其对应的、定义在 n 个顶点上的图的概率分布 $\mathcal{D}$ 。那么，存在一个定义在 n 个顶点上的着色模型，使得每个顶点最多有 $2^{\binom{n}{2}} + 1$ 种颜色，并产生相同的分布 $\mathcal{D}$。

1.3 开放问题

仍有许多有趣的问题尚未解决。最明显的是，对于较大的 n ，匹配独立性和边子图独立性的上界与下界之间存在间隔。对于着色模型，当颜色数量无界以及对于 $k = 2$时，我们确实有相匹配的上界和下界，但对于 k的其他常数值则没有。此外，对于固定的 n ，尚不清楚通过增加更多颜色，着色模型能丰富到何种程度。如上所述，我们知道 $\mathcal{G}_{\mathrm{col}}(n,p) = \mathcal{G}_{\mathrm{col},\ k}(n,p)$与 $k = 2^{\binom{n}{2}} + 1$ 以及 $\rho_{\mathrm{col}}(n) = \rho_{\mathrm{col},\ n-1}(n)$。但仍未解决的问题是函数 $k^{\mathcal{G}}_{min}(n) = \min\{k \mid \mathcal{G}_{\mathrm{col}}(n,p) = \mathcal{G}_{\mathrm{col},\ k}(n,p)\}$ 以及$k^{\rho}_{min}(n) = \min\{k \mid \rho_{\mathrm{col}}(n,p) = \rho_{\mathrm{col},\ k}(n,p)\}$的行为，也就是说，分别需要多少种颜色，才能表达 n 个顶点上的所有着色模型，以及刻画 n 个顶点上的着色模型的连通性阈值的行为?

[5]事实上，这些结果与密度图兰情形一样，仅针对有限多种颜色进行了表述，但证明在我们更一般的设定下同样成立。

最后，本文聚焦于连通性，因为可以说这是图最基本的全局性质。我们仅将此视为一个起点，并且认为探究其他全局性质的类似问题会很有意义。一个自然的拓展方向是研究能够以非零概率实现的最大连通分量的大小，但对于其他任何全局图性质也会产生同样的问题，例如哈密顿性、色数，以及许多其他性质。

1.4 证明技巧

我们通过给出具体的分布示例来证明所考虑的独立性条件具有严格的层级结构（定理 6）：这些分布满足较弱的独立性条件，但不满足较强的条件。所有这些示例都易于描述，且颇具说明性。

阈值 ρ 的下界通过一系列定义在非连通图上的图分布给出。这里，ρ_{pw}（成对独立性）的下界使用的分布，与用于证明 $\mathcal{G}_{\mathrm{pw}} \neq \mathcal{G}_{\mathrm{mat}}$ 的分布相同。ρ_{mat}、ρ_{esub} 和 ρ_{sub} 的下界都使用相同的构造，因为我们无法利用边子图独立性乃至匹配独立性条件所提供的额外自由度（除非在匹配独立性模型中，n 可被 8 整除）。ρ_{sub} 的这个下界此前已在 [15，定理 16] 中得到证明。为完整起见，我们给出该构造。它使用的分布在概念上可视为一种具有复概率的着色模型。该分布满足子图独立性的证明，基于着色模型具有子图独立性这一事实以及代数学基本定理。ρ_{col} 和 ρ_{col_2} 的下界同样来自简单的构造。

对于阈值 ρ 的上界，我们根据独立性条件采用截然不同的证明技巧。对于成对独立性和匹配独立性，上界的证明并未利用这些独立性条件，而只是结合期望的线性性论证与非连通图中边数的最大值。如上所述，边子图独立性是第一个使我们能够应用 Lov´asz 局部引理，从而为 ρ得到常数上界的独立性条件。对于着色模型，我们给出具体策略，说明如何为每个顶点选取一种颜色，使图保持连通。对于两种颜色的情形，这一策略相当简单；而对于一般情形，该策略则基于调整随机着色，该着色在 n足够大时以高概率满足一些有用的性质。这个上界的证明（定理 32）无疑是本文技术性最强的证明。

1.5 论文概述

在第 2 节中，我们介绍证明中用到的一些概率论经典工具。接着，在第 3 节中，我们证明定理 6，即我们的五种独立性模型构成一个严格层级。第 4 节包含定理 7 中总结的所有上界和下界的证明。

2 预备知识

我们使用以下版本的 Lov´asz 局部引理。

引理 9 ([30])。设 $A_1, \dots A_k$ 是一列事件，且每个事件发生的概率至少为 p，并且除至多 d 个其他事件外，与所有其他事件相互独立。那么，如果

$$p > \begin{cases} 1 - \frac{(d-1)^{d-1}}{d^d} & for\ d \geq 2, \\ 1/2 & for\ d = 1, \end{cases}$$

所有事件都发生的概率非零。

下述引理可直接由马尔可夫不等式推出。

Lem引理 10. 设 X 是满足 $Pr[X \leq u] = 1$的随机变量。那么对于 $\ell < u$,

$$Pr[X \geq \ell] \geq \frac{\mathbb{E}[X] - \ell}{u - \ell}.$$

证明。 注意，随机变量 $u - X$ 非负，且其期望为 $u - \mathbb{E}[X]$。因此，我们可以应用马尔可夫不等式。

$$Pr[X \geq \ell] = Pr[u - X \leq u - \ell] = 1 - Pr[u - X > u - \ell] \geq 1 - \frac{u - \mathbb{E}[X]}{u - \ell} = \frac{\mathbb{E}[X] - \ell}{u - \ell}. \qquad \square$$

我们还会用到阿祖马不等式，例如 [21] 中所述的形式；该不等式最初由 Azuma [4]证明：

定理 11. 设 $(\Omega\ Pr)$ 是由 N个离散概率空间 $\{(\Omega_i, Pr_i)\}_{1 \leq i \leq N}$的乘积构成的概率空间，并设 $X\colon \Omega \to \mathbb{R}$ 是一个随机变量，其第 i个坐标的影响至多为 c_i（即，对于每个 $j \neq i$ 以及所有 $\omega_i, \omega_i' \in \Omega_i$，对所有 $\omega_j \in \Omega j$，有 $|X(\omega_1, \ldots, \omega_i, \ldots \omega_N) - X(\omega_1, \ldots, \omega_i', \ldots, \omega_N)| \leq c_i$）。那么，对于所有 $t \geq 0$，有

$$Pr[X \geq \mathbb{E}[X] + t] \leq e^{-\frac{t^2}{2\sum_{i=1}^N c_i^2}} \ and \ Pr[X \leq \mathbb{E}[X] - t] \leq e^{-\frac{t^2}{2\sum_{i=1}^N c_i^2}}.$$

在我们的一些证明中，我们使用以下方法从两个图分布构造一个新的图分布。

定义 12. 设 C, D 为定义在 n 个顶点的图上的两个分布。对于任意 $0 \leq \alpha \leq 1$，其凸组合 $\alpha \cdot C + (1 - \alpha) \cdot D$ 是通过以下方式得到的分布：首先抛掷一枚正面朝上的概率为 α的偏置硬币，然后若硬币正面朝上，则从 C 中采样；否则，从 D 中采样。

引理 13. 设 C, D 是定义在 n 个顶点上的图的两个分布，且对于每条边 e， e 存在的边缘概率在两个分布中相同，即 $Pr_C[X_e] = Pr_D[X_e]$。那么，对于任意 α，如果 C 和 D 分别是两两独立、匹配独立或边子图独立的，则 $\alpha \cdot C + (1 - \alpha) \cdot D$ 也分别具有相同性质。

证明可见附录 B。注意，条件 $Pr_C[X_e] = Pr_D[X_e]$ 不可省略。没有这个条件，凸组合仍然是概率分布，但独立性会丧失。要看出这一点，只需注意，n 个顶点的图上的每个单点分布（即对于固定图 G_0 的 $\mathbf{Pr}[G = G_0] = 1$ ）都显然是着色模型，因此也满足所有其他独立性条件。此外，n 个顶点的图上的每个 分布都可以写成单点分布的凸组合；若引理在没有条件 $Pr_C[X_e] = Pr_D[X_e]$的情况下成立，就会由此得出矛盾。另请注意，该引理不适用于子图独立性，也不适用于着色模型；在引理 15 的证明中会给出一个反例。

3 严格层级

本节旨在证明定理 6:

定理 6. $\mathcal{G}_{col} \subsetneq \mathcal{G}_{sub} \subsetneq \mathcal{G}_{esub} \subsetneq \mathcal{G}_{mat} \subsetneq \mathcal{G}_{pw}$。

我们将从左到右，分别用一个引理证明每个严格包含关系。这些引理中的第一个表明，每个着色模型都是子图独立的，但有些子图独立分布并非着色模型；这一点已在先前的研究中得知 [1, 11, 15, 20], ，不过为了完整起见，我们仍在此给出证明。

引理 14. $\mathcal{G}_{col} \subsetneq \mathcal{G}_{sub}$，也就是说，每个着色模型都是子图独立的，但存在一些并非着色模型的子图独立分布。

证明。 我们先证明 $\mathcal{G}_{\text{col}} \subseteq \mathcal{G}_{\text{sub}}$。考虑一个图分布 $D \in \mathcal{G}_{\text{col}}$，以及任意两个互不相交的顶点子集 V，W。V 中生成的图仅取决于 V 上的基本实验，而 W 中生成的图仅取决于 W 上的基本实验。由于这些基本实验彼此独立，因此 V 和 W 中生成的图相互独立。由此， $D \in \mathcal{G}_{\text{sub}}$。

接下来，我们证明 $\mathcal{G}_{\text{col}} \neq \mathcal{G}_{\text{sub}}$。[6] 考虑图分布 $\mathcal{T}$，它从 3 个顶点且含 1 条边的三个图中均匀选取一个。由于只有三个顶点， $\mathcal{T}$ 显然是子图独立的。为导出矛盾，我们假设它同时也是一个着色模型。考虑任意顶点 v 、其逆时针邻居 u 和顺时针邻居 w。在 v 处进行的基本实验的结果可以分解为两类：允许边 $\{u, v\}$（但不允许 $\{w, v\}$）存在的结果，以及允许边 $\{w, v\}$（但不允许 $\{u, v\}$）存在的结果。不可能有其他结果，因为否则所得图将以非零概率含有零条边或多于一条边，而这在 $\mathcal{T}$ 中并不会发生。现在考虑每个顶点处实验的结果都属于允许其与顺时针邻居之间的边存在（且不允许其与逆时针邻居之间的边存在）的那一部分的情形。该情形必然以非零概率发生，并会得到一个不含边的图。这与 $\mathcal{T}$ 的定义矛盾，因此我们得出 $\mathcal{T}$ 不是一个着色模型。于是，$\mathcal{G}_{\text{col}} \neq \mathcal{G}_{\text{sub}}$。

请注意， $\mathcal{T}$ 也可能作为诱导子图出现在更大的子图独立图分布中，因此，即使仅考虑定义在某个固定数量 $n > 2$ 个顶点上的图分布，该包含关系仍然是严格的。

□

引理 15. $\mathcal{G}_{sub} \subsetneq \mathcal{G}_{esub}$，即每个子图独立分布都是边-子图独立的，但存在一些边-子图独立分布并非子图独立分布。

证明。 显然， $\mathcal{G}_{\text{sub}} \subseteq \mathcal{G}_{\text{esub}}$，根据定义，子图独立性意味着边子图独立性，因为单条边也是一个子图。

为证明 $\mathcal{G}_{\text{sub}} \neq \mathcal{G}_{\text{esub}}$，我们考虑定义在 $n \geq 6$ 个顶点上的如下图分布 $CC(n)$：以 $1/2$ 的概率，该分布返回一个从 Erd˝os-R´enyi 分布$G_{n\text{中}1/2}$抽取的图。否则，该分布会均匀随机地将顶点集 $[n]$分解为两个集合 A 和 B，并返回由 A 和 B上的团的并集构成的图。我们首先证明 $CC(n)$ 是边子图独立的。无论是在 Erd˝os-R´enyi 情形还是在双团情形下，对于任意边 e，都有 $Pr[X_e] = 1/2$。即使在给定任意一个与 e不相交的子图内的结果后，这一结论仍然成立，因此 $CC(n)$ 是边子图独立的。

[6]请注意，这个不等式实际上可以由定理 7 中的界直接推出。尽管如此，为了帮助进一步理解，我们仍然给出一个子图独立分布的示例，该分布并非着色模型。

另一方面，我们证明 $CC(n)$ 不是子图独立的。为此，我们将顶点集 $[n]$ 分解为集合 $P = \{u, v, w\}$ 和 $Q = [n] \setminus P$。我们考虑两个事件 $\mathcal{P}$:= "P 内恰有两条边 " 和 $\mathcal{Q}$:= "Q 是一个团 "。显然，我们有

$$Pr[\mathcal{Q}] = \frac{1}{2} \cdot 2^{-(n-4)} + \frac{1}{2} \cdot 2^{-\binom{n-3}{2}} > 2^{-\binom{n-3}{2}}.$$

另一方面， $\mathcal{P}$ 意味着我们处于 Erd˝os-R´enyi 模型范围内，因此

$$Pr[\mathcal{Q}|\mathcal{P}] = 2^{-\binom{n-3}{2}}.$$

我们得出结论， $Pr[\mathcal{Q}] > Pr[\mathcal{Q}|\mathcal{P}]$ 因此 $\mathcal{P}$和 $\mathcal{Q}$并不独立，这表明 $CC(n)$不是子图独立。

□

我们在上述证明中构造的图分布 $CC(n)$是其两个状态的凸组合，如定义 12 所定义。

引理 16。 $\mathcal{G}_{esub} \subsetneq \mathcal{G}_{mat}$，即每个边-子图独立分布都是匹配独立分布，但存在一些匹配独立分布并非边-子图独立分布。

证明。 为证明 $\mathcal{G}_{\text{esub}} \subseteq \mathcal{G}_{\text{mat}}$，考虑某个边-子图独立分布。为证明它也是匹配独立的，设 $M = \{e_1, \ldots, e_k\}$ 为某个匹配。我们证明对所有 $i \leq k$，都有 $Pr[X_{e_i}|X_{e_1}, \ldots, X_{e_{i-1}}] = Pr[X_{e_i}]$。这意味着

$$Pr[X_{e_1} \text{ and } X_{e_2} \text{ and } \ldots \text{ and } X_{e_k}] = \prod_{i=1}^{k} Pr[X_{e_i}],$$

从而推出匹配独立性。令 V' 为边 e_1、.....、 e_{i-1} 的端点集合。此外，令 $\mathcal{G}(V')$ 为 V' 上所有可能的图的集合。于是，根据全概率公式，我们有

$$Pr[X_{e_i}|X_{e_1}, \ldots, X_{e_{i-1}}] = \sum_{G \in \mathcal{G}(V')} Pr[X_{e_i}|G] Pr[G|X_{e_1}, \ldots, X_{e_{i-1}}].$$

由于边子图独立性， $Pr[X_{e_i}|G] = Pr[X_{e_i}]$，因此我们有

$$Pr[X_{e_i}|X_{e_1}, \ldots, X_{e_{i-1}}] = Pr[X_{e_i}] \sum_{G \in \mathcal{G}(V')} Pr[G|X_{e_1}, \ldots, X_{e_{i-1}}] = Pr[X_{e_i}],$$

从而证明了所需的结论。

对于该陈述的第二部分， $\mathcal{G}_{\text{esub}} \neq \mathcal{G}_{\text{mat}}$，我们考虑如下的图分布$SC(n)$：对于奇数 $n \geq 5$，在 n 个顶点上，每个顶点都独立且均匀地从红色和蓝色这两种颜色中选择一种。由于 n 为奇数，其中一种颜色必然被偶数个顶点选中。具有该颜色的顶点构成一个团，且不存在其他边。

我们首先证明 $SC(n)$ 是匹配独立的：每条边单独出现的概率为 $1/4$，因为首先，两个端点必须具有相同的颜色；其次，该颜色必须是被偶数个顶点选中的颜色。我们证明，包含 k 条边的匹配出现的概率为 $(1/4)^k$。要使该匹配出现，所有 $2k$ 个顶点必须选择相同的颜色，这一事件发生的概率为 $(1/2)^{2k-1}$。其次，该颜色必须最终成为被偶数个顶点选中的颜色，其概率为 $1/2$。综上，其概率为 $(1/2)^{2k-1+1} = (1/4)^k$。

最后，我们证明 $SC(n)$ 不具备边子图独立性。设 W 是一组 $n-2$ 顶点， e 是剩余两个顶点之间的边。若 W 是一个团，则 e 不可能存在。由于 W 是一个团以及 e 存在这两种情况的边缘概率均非零，因此边子图独立性不可能成立。

□

引理 17. $\mathcal{G}_{mat} \subsetneq \mathcal{G}_{pw}$，即每个匹配独立分布都是两两独立的，但存在一些两两独立分布并非匹配独立分布。

为证明这个最后的严格包含引理，我们定义如下分布。稍后我们也会用这个分布来证明 ρ_{pw} 的一个下界。

定义 18. $CM(n,q)$ 对于 n 为偶数时，是定义在 n 个顶点的图上的如下分布：若要采样 $G \sim CM(n,q)$，以 q 的概率从团情形中采样，否则从匹配情形中采样。在团情形中，我们均匀随机地选取一个顶点 $x \in [n]$，并添加所有与 $x \notin e$ 相连的边 e，其中 $x \notin e$ 至 G。在匹配情形中，我们均匀随机地选取 $[n]$ 上的一个完美匹配，并将其边添加到 G 上。

断言 19. 存在 $0 \leq q(n) \leq 1$，满足

$$q(n) = 1 - \Theta\Big(\frac{1}{n^2}\Big)$$

使得，对于 n 为偶数，$CM(n, q(n))$ 是两两独立的，但不是匹配独立的，并且每条边的概率为

$$p(n) = 1 - \frac{2}{n} - \Theta\Big(\frac{1}{n^2}\Big).$$

这个断言的证明很直接，只需进行一些计算，因此我们将其推迟到附录 C。

引理 *17* 的证明。显然，$\mathcal{G}_{\text{mat}} \subseteq \mathcal{G}_{\text{pw}}$ 因为根据定义，匹配独立性蕴含两两独立性；两个顶点不相交的边 e，f 构成一个匹配。

为说明 $\mathcal{G}_{\text{mat}} \neq \mathcal{G}_{\text{pw}}$，回忆一下，根据断言 19，存在某个 $q(n)$，使得 $CM(n,\ q)$ 两两独立，但并非匹配独立。 □

定理 *6* 的证明。定理 6 现在可直接由引理 14 至 17 得出。 □

4 关于 ρ的界

本节将证明定理 7，给出我们各种独立性模型下 ρ 的下界和上界。注意，根据第 1.2 节中 ρ 的定义，对于某个独立性条件 X，ρ_X 的任何下界也适用于某个较弱的独立性条件 Y 下的 ρ_Y，也就是说，该条件满足 $\mathcal{G}_{\text{X}} \subseteq \mathcal{G}_{\text{Y}}$。反过来，$\rho_Y$ 的任何上界也适用于 ρ_X。

4.1 两两独立性

引理 20. 对于偶数 $n \geq 4$，有 $\rho_{pw}(n) \geq 1 - \frac{2}{n} - \Theta(\frac{1}{n^2})$。 —

证明。考虑定义 18 中定义的分布 $CM(n,q)$，选取 $q = 1 - \Theta\Big(\frac{1}{n^2}\Big)$ 使得 $CM(n,q)$ 两两独立；根据断言 19，这是可以做到的。在 $CM(n,q)$ 中，边的边缘概率 p 因而为 $p = 1 - \frac{2}{n} - \Theta\Big(\frac{1}{n^2}\Big)$。注意，$G \sim CM(n,q)$ 连通的概率为 0，因为在团情形中总会有一个孤立顶点，而在匹配情形中，G 仅由一个完美匹配构成。因此，两两独立模型中连通性具有非零概率的阈值 ρ_{pw} 至少为 $p = 1 - \frac{2}{n} - \Theta\Big(\frac{1}{n^2}\Big)$。

– — □

引理 21. 对于任意 n，都有 $\rho_{pw}(n) \leq 1 - \frac{2}{n}$。 –

证明. 如果任意图分布（甚至不一定是两两独立的）中的最小边边缘概率 p 大于 $1-\frac{2}{n}$，那么边数的期望 $\mathbb{E}[\sum_e X_e]$ 大于 $(1-\frac{2}{n})\binom{n}{2} = \binom{n-1}{2}$。由于任何非连通图至多包含 $\binom{n-1}{2}$ 条边，因此该图以非零概率为连通图。

□

4.2 匹配独立性

目前没有专门针对匹配独立性的通用界。匹配独立性这一条件出人意料地难以利用。因此，我们给出以下推论，它可直接由引理 21 和引理 26 得出。

推论 22. 对于任意 $n \geq 2$，我们有 $\frac{1}{2}(1-\tan^2\frac{\pi}{2n}) \leq \rho_{mat}(n) \leq 1-\frac{2}{n}$。

当 n 能被 8 整除时，推论 22 中的下界可以略微改进；这尤其表明，对于无穷多个 n，$\rho_{mat}(n) \neq \rho_{sub}(n)$，再次说明不同独立性条件下的连通性阈值表现各异。

引理 23. 对于任意能被 8整除的 $n \in \mathbb{N}$，我们有 $\rho_{mat}(n) \geq \frac{1}{2}$。

证明. 我们通过给出一种基于顶点着色的匹配独立分布来证明这一点。考虑所有可能的红蓝顶点着色 $c\colon [n] \to \{red, blue\}$，并为其赋予权重函数 $w(c) = a_{red(c)}2^{-n}$，其中 $red(c)$ 是着色c 中红色顶点的数量，并且

$$a_k = \begin{cases} 1, & \text{if } k \equiv 1 \mod 2, \\ 2, & \text{if } k \equiv 2 \mod 4, \\ 0, & \text{if } k \equiv 0 \mod 4. \end{cases}$$

注意到

$$a_k = 1^k - \frac{1}{2}i^k - \frac{1}{2}(-i)^k,$$

其中 i 是虚数单位，我们得到的是

$$\sum_{\text{all colorings } c} w(c) = 2^{-n} \sum_{k\in\{0\}\cup[n]} a_k \binom{n}{k} = 2^{-n}\Big((1+1)^n - \frac{1}{2}(1+i)^n - \frac{1}{2}(1-i)^n\Big)$$

通过二项式展开。注意，$(1+i)^4 = (2i)^2 = -4$ 和$(1-i)^4 = (-2i)^2 = -4$，因此当 n 能被 8 整除时，我们有

$$\sum_{\text{all colorings } c} w(c) = 2^{-n}\Big(2^n - (-4)^{n/4}\Big) = 1 - 2^{-n/2} < 1.$$

因此，我们可以定义一个关于着色的概率分布：以概率 $w(c)$ 按照着色 c 对顶点进行着色，以概率 $2^{-n/2}$ 使所有顶点都保持未着色。随后，我们通过在所有红色顶点上添加一个团，并在所有蓝色顶点上添加一个团来构造图（如果顶点未着色，则保持图为空）。注意，所得的图永远不会是连通的，因为 $a_n = a_0 = 0$，这意味着不可能所有顶点都是蓝色，也不可能所有顶点都是红色。

我们现在证明该分布是匹配独立的。也就是说，对于 K_n 上的任意非空匹配 M，我们证明，如果 G 是从上述分布中采样得到的，那么 $Pr[M \in G] = 2^{-|M|}$。令 $e \in M$，并考虑不属于 e 的顶点的着色，这些着色满足了以下条件。

$M \setminus \{e\}$ 中的边存在于 G 中。注意，要使 $f \in M \setminus e$ 存在，我们需要 f 中的两个顶点具有相同的颜色。因此，对于不在 e 中的顶点，共有 $2^{n-1-|M|}$ 种不同的着色方式，使得 $M \setminus e$ 中的所有边都存在。设 c 为其中一种（部分）着色。如果 $red(c)$ 为奇数，则有两种方式可以完成着色，使得 e 存在（其两个端点要么均为红色，要么均为蓝色），并且每种最终着色出现的概率均为 2^{-n}。如果 $red(c) \equiv 0 \mod 4$，那么要使 e 存在且该着色以正的（$2 \cdot 2^{-n}$）概率出现，完成这种着色的唯一方式是将 e 的两个端点都着为红色。类似地，如果 $red(c) \equiv 2 \mod 4$，那么要使 e 存在且该着色以正的（$2 \cdot 2^{-n}$）概率出现，完成这种着色的唯一方式是将 e 的两个端点都着为蓝色。因此，

$$Pr[M \in G] = 2^{n-1-|M|} \cdot 2 \cdot 2^{-n} = 2^{-|M|},$$

如所需。 □

4.3 边子图独立性

我们没有针对边子图独立性的下界。这个下界继承自更严格的子图独立性，即引理 26。

推论24。 对于任意 $n \geq 2$，我们有 $\rho_{esub}(n) \geq \frac{1}{2}(1 - \text{正切}^2 \frac{\pi}{2n})$。

引理25。 对于任意 n，我们有 $\rho_{esub}(n) \leq \frac{3}{4}$。

证明。 我们证明，任何最小边边缘概率为 $p > \frac{3}{4}$ 的边-子图独立分布都有非零概率是连通的。为此，我们使用引理 9 中所述的 Lov´asz 局部引理。

我们选取任意一条哈密顿路径 e_1、.....、e_{n-1} 中的边，并考虑与之对应的事件 X_{e_1}、.....、$X_{e_{n-1}}$。由于具有边子图独立性，这些事件中的每一个至多依赖于其他事件中的 $d = 2$ 个（即与相邻边对应的事件）。此外，每个事件的概率至少为 $p > \frac{3}{4}$。根据引理 9，由于 $p > \frac{3}{4} = 1 - \frac{1}{2^2}$，所有事件 X_{e_1}、.....、$X_{e_{n-1}}$ 以非零概率同时发生，此时哈密顿路径的所有边都存在，因此该图是连通的。 □

4.4 子图独立性

ρ_{sub} 的下界已在 [15, 定理 16], 中给出，但为完整起见，我们在此提供证明。

引理 26 （[15, 定理 16]）。 对于任意 $n \geq 2$，我们有 $\rho_{sub}(n) \geq \frac{1}{2}(1 - \text{正切}^2 \frac{\pi}{2n})$。

为了证明这个引理，我们在 $n \geq 2$ 个顶点上构造一个子图独立的分布，使每条边出现的边缘概率等于 $p = \frac{1}{2}(1 - \tan^2 \frac{\pi}{2n})$，且从该分布中抽样得到的图为连通图的概率为 0。

令 $\mathcal{K}_n$ 为顶点集 $[n]$ 上所有至多有两个连通分量且每个连通分量都是团的图组成的集合。我们将在 $\mathcal{K}_n$ 上构造一个概率分布，其支撑集包含所有这类图，但不包括 K_n，并且满足子图独立性的条件。

令 $z = \frac{1}{2} + \frac{i}{2} \tan \frac{\pi}{2n}$。注意， z 是一个复数。对于任意图 $G \in \mathcal{K}_n$，我们定义

$$p_G := z^a(1-z)^{n-a} + (1-z)^a z^{n-a}, \tag{1}$$

其中， a 表示 G中最大团的大小。

注意到，将 z 和 $1-z$ 写成极坐标形式后，可以用三角不等式改写。我们有 $z = \frac{e^{\pi i/2n}}{2}\cos \pi/2n$ 和 $1-z = \frac{e^{-\pi i/2n}}{2}\cos \pi/2n$，因此有

$$p_G = \frac{1}{2^n \cos^n \pi/2n}\left(e^{\pi i(a/n-1/2)} + e^{-\pi i(a/n-1/2)}\right) = \frac{\cos \pi(a/n-1/2)}{2^{n-1}\cos^n \pi/2n}, \tag{2}$$

由此我们立即可知，p_G 是非负实数（因为 $1/2 \le a/n \le 1$），且满足 $p_{K_n} = 0$（因为 $a = n$ 蕴含 $a/n - 1/2 = 1/2$，并且 $\cos \pi/2 = 0$）。

命题 27. 数值 p_G 在 $\mathcal{K}_n$ 上定义了一个概率分布，其边密度为$z^2 + (1-z)^2 = \frac{1}{2}(1-\tan^2 \frac{\pi}{2n})$，该分布具有子图独立性，且图不连通的概率为1。

证明。 我们已经看到，对于所有 $G \in \mathcal{K}_n$，$p_G \ge 0$ 成立；而对于唯一的连通图，即 K_n，p_G 为 0。因此，要证明这些数值定义了一个概率分布，剩下只需证明

$$\left(\sum_{G\in\mathcal{K}_n} p_G\right) - 1 = 0.$$

为了证明 G 的边密度为 $z^2 + (1-z)^2$,，我们必须证明对所有 $e \in \left(\frac{[n}{2}\right]$

$$\left(\sum_{G\ni e} p_G\right) - z^2 - (1-z)^2 = 0.$$

此外，假设满足概率分布的条件，子图独立性等价于：对于所有不相交顶点集 $U, V \subseteq [n]$，以及分别定义在 U 和 V 上的所有图 H_1, H_2，我们有

$$\left(\sum_{G[U]=H_1 \text{ and } G[V]=H_2} p_G\right) - \left(\sum_{G[U]=H_1} p_G\right)\left(\sum_{G[V]=H_2} p_G\right) = 0.$$

由于这些值 p_G 是关于 z 的多项式，因此这些条件中的每一个都是关于 z 的多项式方程，其值必须为 0。由于我们关注的是 $z = \frac{1}{2} + \frac{i}{2}\tan \frac{\pi}{2n}$ 的情形，因此需要证明它们在该值下取值为 0。事实上，我们证明它们对所有 z 都取值为 0，即它们都等于零多项式。为证明这一点，我们证明这些条件对 $z \in [0,1]$ 均成立。

任取 $z \in [0,1]$。考虑 $[n]$ 上的随机图分布：将每个顶点以 z 的概率着为红色，以 $1-z$ 的概率着为蓝色；当且仅当两个顶点颜色相同时，才在它们之间连一条边。现在注意到，(1) 恰好定义了得到任意图 $G \in \mathcal{K}_n$ 的概率，因此对于这个 z 的取值，p_G 的各个值构成一个边密度为 $z^2 + (1-z)^2$ 的概率分布。此外，由于该分布是一个着色模型，因此它也是子图独立的。因此，对于这一 z 的取值，前面提到的所有多项式都为零。

由于 z 中唯一具有无穷多个零点的多项式是零多项式，因此这些多项式对任意复数 z 的取值都必须为零，结论由此得证。 □

引理 *26* 的证明。 根据命题 27，式 (1) 中定义的概率构成了一个定义在具有 n 个顶点的图上的子图独立概率分布，其中每条边的边缘概率恰为 $\frac{1}{2}(1-\tan^2 \frac{\pi}{2n})$，且图连通的概率为零。因此，$\rho_{\text{sub}}(n) \ge \frac{1}{2}(1-\tan^2 \frac{\pi}{2n})$。 □

关于 ρ_{sub} 的匹配上界也可以在 [15]，中找到，不过我们在此不再重述证明。

引理 28 ([15,定理 16])。 对于任意 $n \ge 2$，我们有 $\rho_{sub}(n) \le \frac{1}{2}(1-\text{正切}^2 \frac{\pi}{2n})$。

4.5 着色模型

我们首先考虑只有两种颜色的着色模型，因为对于这种情况，我们可以找到匹配的下界和上界。着色模型的其他限制条件也值得研究，例如颜色数量有其他上限的情况，或者每个顶点都必须从其颜色中均匀地 选择一种颜色的情况。

4.5.1 两种颜色

我们证明，在这种情况下，阈值概率恰好为 $1/4$。

引理 29. 对于所有 $n \geq 3$，我们有 $\rho_{col,2}(n) = 1/4$。

证明。我们首先通过定义以下始终不连通的着色模型来证明$\rho_{col,2}(n) \geq 1/4$。将顶点集分解为三个非空集合 A、B、C。每个顶点都从红色和蓝色这两种颜色中均匀随机地选择一种。同一集合中的两个顶点在它们都为红色时连通。若 a 为红色且 b 为蓝色，则顶点 $a \in A$ 与 $b \in B$连通。对于 $b \in B$ 和 $c \in C$，以及 $c \in C$ 和 $a \in A$，情况也相同。显然，每条边的边缘概率均为 $1/4$。现在我们证明，从该分布中采样得到的图的每个连通分量，都是集合 A、B、C中至多两个集合的并集的子集。为此，我们选取一条边 $\{a, b\}$，使得 $a \in A$ 和 $b \in B$ 均存在，并尝试扩展其连通分量。由于边 $\{a, b\}$存在，因此 a 必为红色，而 b 必为蓝色。现在，a 只能与 A中的其他红色顶点以及 B中的蓝色顶点连通。类似地，b 只能与 A中的其他红色顶点连通。A 中的红色顶点或 B 中的蓝色顶点均无法与 C中的顶点连通，因此，包含边 $\{a, b\}$ 的连通分量包含于 $A \cup B$之中。对任意一对集合，利用对称性也可得到相同结论，由此我们得出该图必定不连通，从而证明了下界。

接下来，我们通过证明 $\rho_{col,2}(n) \leq 1/4$ 来说明：如果每条边的边缘概率都严格大于 $1/4$，我们总能找到一种能够使图连通的着色。若对于顶点 v**覆盖** 顶点 w，则称：对于顶点 w的两种颜色，在 v处都存在一种颜色，使得边 $\{v, w\}$在该着色下存在。现在，我们逐个移除顶点：反复移除一个覆盖某个剩余顶点的顶点，直到不存在这样的顶点为止。如果剩余顶点所构成的图 V' 可以通过某种着色实现连通，我们就可以按相反顺序加回被移除的顶点，并将它们连接到其所覆盖的顶点，从而使整个图连通。因此，我们只需证明图 V' 可以连通。由于在该图中没有顶点覆盖其他顶点，每条边只会在其两个端点的四种可能颜色组合中的恰好一种组合下存在。

我们选取一个顶点 $v' \in V'$，使其在所有 $v \in V'$中最大化 $\max(Pr[v\text{为红色}], Pr[v\text{为蓝色}])$。我们为该顶点 v' 赋予概率更大的颜色，令 p 为该颜色的概率。可以看出，对于任意其他顶点 $w \in V' \setminus \{v'\}$，必然存在一种颜色，使得边 $\{v', w\}$存在。否则，使 $\{v', w\}$ 存在的颜色组合的概率至多为 $p \cdot (1 - p)$，而这不超过 $1/4$。因此，只需为每个顶点赋予能够与 v'连接的正确颜色，我们就能将 V' 连接起来。上界的证明至此完成，整个引理也随之得证。

□

4.5.2 一般着色模型

现在，我们考虑着色模型的完整一般性：每个顶点处的局部实验可以有无界甚至无限多个结果（不过根据引理 8，不失一般性地可以排除这种情况）。

引理 30. 对于所有 $n \geq 3$，我们有 $\rho_{col}(n) \geq (n-2)^{\frac{3n-4-\sqrt{5n^2-16n+12}}{2(n-1)^2}}$。

注意，对于 $n = 3$，该界等于 $1/4$（这与上一小节相符，因为在这种情况下，我们的构造只使用两种颜色）；而对于 $n \to \infty$，它趋近于 $2 - \phi \approx 0.381966$。我们指出，引理 30 是在 [5, 定理 1.7]中同时且独立地得到的，且使用了完全相同的构造。

证明。我们证明，对于每个顶点数 $n \geq 3$，都存在一个边的边缘概率为 $p(n) = (n-2)^{\frac{3n-4-\sqrt{5n^2-16n+12}}{2(n-1)^2}}$的着色模型，使其以概率 0 连通。

对于任意 $n \geq 3$，我们定义分布 $CS(n)$。固定顶点 v 将另一个顶点 $v' \in [n] \setminus \{v\}$ 均匀随机地选作其局部实验。所有其他顶点以概率 q（稍后确定）选择红色，以概率 $1-q$选择蓝色。随后，v 连接到 $V \setminus \{v, v'\}$ 中所有着蓝色的顶点。在 $V \setminus \{v\}$中的顶点之间，当且仅当两个端点都着红色时，才存在一条边。所得的图显然不连通：每个红色顶点只与其他红色顶点相连；而当 $V \setminus \{v\}$ 中的所有顶点都是蓝色时，仅存在 $n-2$条边。

对于每条不与 v 关联的边 e，我们有 $Pr[X_e] = q^2$；对于每条与 v 关联的边 e'，我们有 $Pr[X_{e'}] = (1-q)^{n-2}$。我们选取使这些概率中的最小值最大的 q。由于当 $q \geq 0$ 时，q^2 随 q 增大，而 $(1-q)^{n-2}$ 随 q 减小，因此当$q^2 = (1-q)^{n-2}$时，最小值达到最大。于是我们可以求解

$$q^2 + \frac{n-2}{n-1}q - \frac{n-2}{n-1} = 0$$

得到

$$q = -\frac{n-2}{2(n-1)} \pm \sqrt{\frac{(n-2)^2}{4(n-1)^2} + \frac{n-2}{n-1}}$$
$$q = \frac{2-n \pm \sqrt{(n-2)^2 + 4(n-1)(n-2)}}{2(n-1)}$$
$$q = \frac{2-n \pm \sqrt{5n^2-16n+12}}{2(n-1)}.$$

这些解中只有一个满足 $q > 0$，即“+”对应的解。由于边的边缘概率 $p(n)$ 等于 $(1-q)^{n-2}$，我们便得到所述上界。 □

现在我们来考虑 $\rho_{\text{col}}(n)$ 的上界。首先给出一个适用于所有 n 的较弱常数上界。请注意，[5], 中定理 1.7 的证明 —— 该成果独立且同时发表 —— 采用的构造与下文引理中的构造基本相同，只是他们优化了计算，从而得到 $\frac{1}{2} - \frac{1}{4n-6}$ 的上界。

引理 31. 对于每个 n，我们有 $\rho_{col}(n) \leq \frac{1}{2}$。

证明。我们证明，在最小边边缘概率为 $p \geq 0.5$ 的每个着色模型中，都能找到一个连通结果。设图的顶点为 v_0、……、v_{n-1}。我们考虑 v_0 的所有可能颜色（设有 k 种），以及在给定 v_0 的颜色后，v_0 与其他

[7]为使证明更易于阅读，我们假设随机实验在每个顶点处的结果数量有限，根据引理 8，我们可以作出这一假设。完全相同的证明策略也适用于无限（甚至不可数无限）概率空间。

每个顶点 v_i 之间的边的条件概率。为方便起见，我们将这些概率表示在一个有 k 行、 $n-1$ 列的表格 A 中，其中 $A_{ij} = Pr[X_{v_0v_j}|c\ (v_0)\ = i]$，其中 $c\ (v_0)\ = i$ 表示 v_0 被着成颜色 i。注意，根据全概率公式，对于每一列 j，我们有

$$\sum_{i=1}^{k} A_{ij} Pr[c(v_0) = i] \geq p \geq 0.5.$$

对所有行求和，可得

$$\sum_{j=1}^{n-1}\sum_{i=1}^{k} A_{ij} Pr[c(v_0) = i] \geq \frac{n-1}{2}$$

$$\sum_{i=1}^{k} Pr[c(v_0) = i] \sum_{j=1}^{n-1} A_{ij} \geq \frac{n-1}{2},$$

并且由于 $\sum_{i=1}^{k} Pr[c(v_0)\ = i] = 1$，后一种表述意味着至少存在一行 i，使得

$$\sum_{j=1}^{n-1} A_{ij} \geq \frac{n-1}{2}.$$

固定一个满足此条件的 i，令 x 为该行中满足 $A_{ij} = 0$的条目 j 的数量，令y 为满足 $0 < A_{ij} \leq \frac{1}{2}$ 的条目 j 的数量，并令 z 为满足 $A_{ij} > \frac{1}{2}$ 的条目-j 的数量。那么必然有 $z \geq x$，否则我们就会得到 -

$$\sum_{j=1}^{n-1} A_{ij} \leq x \cdot 0 + y \cdot \frac{1}{2} + z \cdot 1 < (x+z) \cdot \frac{1}{2} + y \cdot \frac{1}{2} = \frac{n-1}{2},$$

这与我们对行 i的选择相矛盾。

为了证明该图以非零概率连通，我们给出一种颜色分配，使得每个顶点都能通过长度至多为 2 的路径到达 v_0。我们为 v_0选择颜色 i。由于我们在上文证明了 $z \geq x$，因此可以将每个满足 $A_{ij} = 0$ 的顶点 v_j 与一个满足 $A_{ij'} > \frac{1}{2}$ 的不同顶点 $v_{j'}$ 配对。鉴于 v_0 的颜色为 i，我们无法直接连接v_j 和 v_0，因此将v_j 连接到$v_{j'}$，并将$v_{j'}$ 连接到 v_0。由于

$$Pr[X_{v_0v_{j'}}|c(v_0) = i] + Pr[X_{v_{j'}v_j}] > \frac{1}{2} + p \geq 1,$$

存在一种颜色选择，使得 $v_{j'}$ 和 v_j 对应的颜色能够使边 $v_0v_{j'}$ 和 $v_{j'}v_j$ 同时存在。我们为 v_j 和 $v_{j'}$ 固定这些颜色。对表格中每个条目为 0 的顶点重复这一过程后，我们只剩下一些满足 $A_{ih} > 0$ 的顶点 v_h （即那些没有与任何条目为 0 的顶点配对的顶点，以及那些表格条目为正数但至多为 $\frac{1}{2}$ 的顶点）。对于每个这样的 v_h，都存在一种颜色选择，使得边 v_0v_h 存在，因此我们选取该颜色。这样便得到了一个以 v_0 为根的生成树，其中所有满足 $A_{ih} > 0$ 的顶点 v_h 都是 v_0 的邻居，而其余顶点都是 v_0 的邻居的邻居（示例见图 3），从而完成证明。

□

接下来，我们给出一个当 n 足够大时成立的更强上界。对于 $n \to \infty$，该上界趋于 $2 - \phi$，因此当 n 足够大时，它与下界相匹配，正如 [5] 中的猜想。

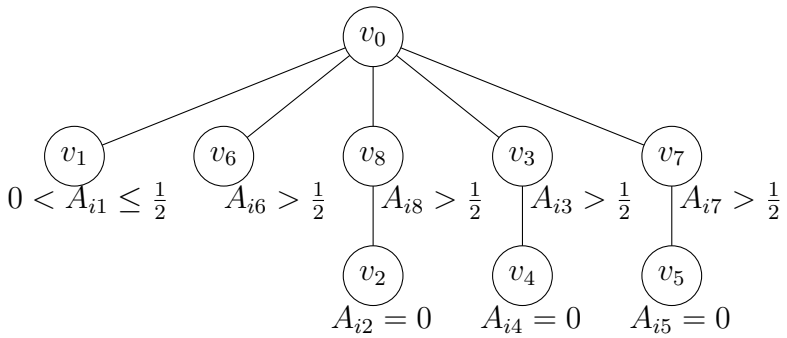


图 3：在此示例中，有三个具有 $A_{ij} = 0$ 的顶点 v_j、四个具有 $A_{ij} > \frac{1}{2}$ 的顶点 v_j，以及一个具有 $0 < A_{ij} \le \frac{1}{2}$ 的顶点 v_j。

定理 32. 对于任意 $\varepsilon > 0$，都存在一个 $n_0(\varepsilon)$，使得对于任意定义在 $n \ge n_0(\varepsilon)$ 个顶点的图上的着色模型分布，只要其最小边边缘概率至少为 $p' = 2 - \phi + \varepsilon$，图以非零概率连通。

为了证明该定理，我们需要以下记号和设定。我们将 p 记作 $2 - \phi$。

观察 33. 对于 $p = 2 - \phi$，我们有以下优美的恒等式：$\frac{1}{p} - 2 = 1 - p$ 和 $(1-p)^2 = p$。

首先，我们按照每个顶点各自的分布，独立地为每个顶点赋予一个随机颜色。[8] 对于这一固定着色，我们用 $N(v, c)$ 表示 v 的邻居集合，其中 v 被重新着色为 c 颜色。

引理 34. 在随机着色中，高概率下，对于所有 v，以下结论同时成立：$\mathbb{E}_c[|N(v,c)|] > (p + \varepsilon/2)n$。

证明。固定一个顶点 v，对于每个 $w \ne v$，令 Y_w 为关于 w 的颜色的随机变量，它表示从 v 的分布中采样颜色时边 vw 存在的概率。注意，$\mathbb{E}[Y_w] \ge p'$。然后

$$Y := \mathbb{E}_c[|N(v,c)|] = \sum_{w \ne v} Y_w.$$

然后，我们可以将 Y_w 作为坐标应用阿祖马不等式。注意，由于 $0 \le Y_w \le 1$，每个 Y_w 对 Y 的影响至多为 1。我们有

$$\mathbb{E}[Y] = \sum_{w \ne v} \mathbb{E}[Y_w] \ge (n-1)p',$$

因此，令 $t := \frac{\varepsilon}{4} n$，利用定理 11，对于足够大的 n，我们得到以下界：

$$Pr[Y \le (p + \varepsilon/2)n] \le Pr[Y \le (p' - \varepsilon/2)n] \le Pr[Y \le \mathbb{E}[Y] - t] \le e^{-\frac{t^2}{2(n-1)}} \le e^{-\varepsilon^2 n/32}.$$

对所有顶点 v 应用并集界可知，事件 $\mathbb{E}_c[|N(v,\ c)|] \le (p + \varepsilon/2)n$ 对于任意顶点 v 发生的概率至多为 $ne^{-\varepsilon^2 n/32}$；当 $n \to \infty$ 时，该概率趋于 0，从而证明了该引理。 □

[8] 我们再次假设颜色数量有限。我们的证明策略也适用于无限概率空间。

因此，从现在起，我们在引理 34 所述的高概率事件条件下进行讨论。这使我们能够论证，可以对具有较大邻域的任意顶点重新着色。此外，利用引理 10，我们还可以论证，对于每个顶点，较大的邻域必定以某个相当大的概率存在。由此，我们可以使用并集界证明，较大的邻域必定会与某些固定的边同时存在。

我们证明图以正概率连通的一般策略是：主要依靠初始时对所有顶点进行的随机着色，但必要时重新着色某些顶点，并利用一些较大的邻域来连接顶点。

由于我们会重新着色某些顶点，最终图中的实际邻域可能与我们据以得到这些上界的随机着色中的邻域不同（也可能更小）。我们用误差项 $C := C'(\varepsilon)\log_2 n$ 作为重新着色的顶点数以及我们试图连接到的顶点数的上界（我们可能会重新着色更多顶点，但并不声称会连接到这些顶点，因此重新着色它们无关紧要）。如果我们能保证随机着色中的某个邻域交集（或某个邻域与一个固定集合的交集）至少包含 $C+1$ 个顶点，就能确定最终实际着色中的交集非空。

令 (r, c_r) 为使 $|N(r, c_r)|$ 最大的顶点-颜色对。我们将 r 重新着色为 c_r。如果每个不属于 $N(r, c_r)$ 的顶点都能重新着色为某种颜色，使其与 $N(r, c_r)$ 中的一个顶点之间有一条边，那么图就能连通。否则，我们从所有满足 $b \notin N(r, c_r) \cup \{r\}$ 且 $|N(r, c_r) \cap N(b, c_b)| \le C$ 的顶点-颜色对中，选取使 $|N(b, c_b)|$ 最大的顶点-颜色对 (b, c_b)。我们将b重新着色为 c_b。

在下文中，我们使用以下简写形式。

$$R := N(r, c_r),\ B := N(b, c_b),\ \alpha := |R|/n,\ \text{and}\ \beta := |B|/n.$$

顶点 r 和 b及其各自的邻域 R 和 B将是我们论证的核心，而我们会通过各种方式将所有其他顶点连接到这些集合。

引理 35. $p < \alpha < 1-p$，且 $p < \beta \le \alpha$。

证明。 我们分别证明这四个界：

$\alpha > p$：由于对所有 v 都有 $\mathbb{E}_c[|N(v,\ c)|] > pn$，显然最大的此类邻域 $\alpha n = |N(r,\ c_r)|$ 也必须大于 pn 。

$\alpha < 1-p$: 如果我们有 $\alpha \ge 1-p$，那么每个其他顶点的一个最大规模邻域（其规模至少为 pn）都必须与 R 相交。我们假设情况并非如此，因此 $\alpha < 1-p$。

$\beta > p$：考虑某个顶点 v，对于每种颜色 v，都有 $(N(v,\ c_v) \cup \{v\}) \cap R = \varnothing$。在求取 $(b,\ c_b)$ 时，所有顶点-颜色对 $(v,\ c')$ 都已纳入考虑。由于 $\mathbb{E}_c[|N(v,\ c)|] > pn$，因此我们也有 $\beta n = |N(b,\ c_b)| > pn$。

$\beta \le \alpha$：这是因为 $(b,\ c_b)$ 是从受限的顶点-颜色对集合中选出的，而 $(r,\ c_r)$ 则是从所有可能的顶点-颜色对中选出的。 □

注意，如果我们有 $R \cap B \neq \varnothing$，就可以轻松地使图连通：如果我们将每个顶点$v \notin R \cup B \cup \{r, b\}$ 重新着成某种颜色 c，使得 $|N(v,\ c)| > pn$，那么 $N(v,\ c)$ 必须与 $R \cup B$ 相交。这是由 $|R \cup B| = |R| + |B| - |R \cap B| \ge \alpha n + \beta n - C \ge 2pn - C$ 和 $(2pn - C) + pn > n$ 得出的。因此，我们在下文中假设 $R \cap B = \varnothing$。

引理 36. $\beta \le \min(1-\alpha, 1/2)$.

证明。 这直接由 $R \cap B = \varnothing$ 得出，因此 $\alpha + \beta \le 1$。 □

为了在剩余情况下使图连通，我们需要引入一些额外的定义。直观地说，如果某个顶点 v 无法与蓝色集合 B稳健地连通，我们就称该顶点为必红。更形式化地说，如果不存在任何颜色 c，使得 $|N(v,c) \cap B| > C$，我们就称顶点 $v \notin B \cup \{b\}$ 为必红顶点。类似地，如果不存在任何颜色 c，使得 $|N(v,c) \cap R| > C$，我们就称顶点 $v \notin R \cup \{r\}$ 为必为蓝色。令 OR 表示必红顶点的集合，令 OB表示必为蓝色顶点的集合。

请注意，没有任何顶点能既必为蓝色又必红。当一个顶点既不是必为蓝色也不是必红时，我们称其为非必然。

我们首先给出关于必为蓝色和必红顶点的邻域大小的一些保证。

引理 37. 设 $u \in OB, v \in OR$。那么，对于 u (v) 的颜色分布， u (v) 的邻域满足以下性质。

1. $Pr[|N(u,c)| \geq \frac{p+p\beta-\beta}{p} n + \varepsilon n/2] \geq 1-p$. *2.* $Pr[|N(v,c)| \geq \frac{p+p\alpha-\alpha}{p} n + \varepsilon n/2] \geq 1-p$. *3.* $Pr[|N(u,c)| \geq (1-p)(1-\beta)n + \varepsilon n/2] \geq 1-\sqrt{p}$. *4.* $Pr[|N(v,c)| \geq (1-p)(1-\alpha)n + \varepsilon n/2] \geq 1-\sqrt{p}$.

证明（概要）。我们可以将任意顶点的邻域大小上界限定为 αn。此外，我们可以将任意必蓝顶点的邻域大小上界限定为 βn。应用引理 10 并化简即可得出该引理。完整计算见附录 D。

□

推论 38. 设 S 为一个大小为 $|S| \leq C$的顶点集。设 $u \in OB$。那么，

$$Pr_c[N(u,c) \cap (B \setminus S) \neq \varnothing] \geq 1-p.$$

证明。 根据引理 37 的第 1 部分，我们有

$$Pr[|N(u,c)| \geq \frac{p+p\beta-\beta}{p} n + \varepsilon n/2] \geq 1-p.$$

我们希望证明，当 $|N(u,c)| \geq \frac{p+p\beta-\beta}{p} n + \varepsilon n/2$时， $N(u,c)$ 与 $B \setminus S$ 相交。由于 $|N(u,c) \cap R| \leq C$ 且 $(B \setminus S) \cap R = \varnothing$，如果这些集合不相交，则 $|N(u,c)| + |B \setminus S|$ 至多为 $n - |R| + C$。因此，只需证明.....

$$\frac{p+p\beta-\beta}{p} n + \varepsilon n/2 + (\beta n - C) \overset{!}{>} n - \alpha n + C,$$

或者等价地，

$$\frac{p+p\beta-\beta}{p} + \varepsilon/2 + \alpha + \beta - 2C/n - 1 \overset{!}{>} 0.$$

首先，我们利用当 n 足够大时， $\varepsilon/4 > 2C/n$ 。

$$\frac{p+p\beta-\beta}{p} + \varepsilon/2 + \alpha + \beta - 2C/n - 1 > \frac{p+p\beta-\beta}{p} + \varepsilon/4 + \alpha + \beta - 1$$

$$= (2 - \frac{1}{p})\beta + \varepsilon/4 + \alpha = (p-1)\beta + \alpha + \varepsilon/4 > (\alpha - \beta) + \varepsilon/4 \geq \varepsilon/4 > 0,$$

由于 $\alpha \geq \beta$。注意，我们在此处使用了观察 33。 □

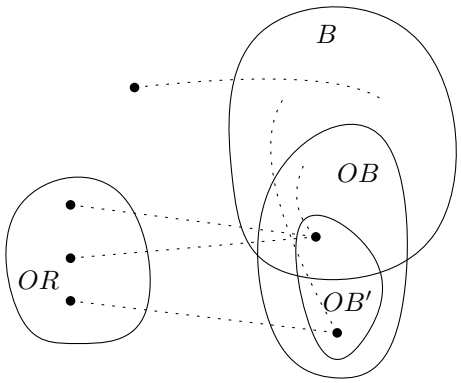


图 4：当 $|OB| > C$ 时如何使图连通。

必红顶点的对称表述同样可由引理 37 得出：

推论 39. 设 S 是大小为 $|S| \leq C$ 的顶点集。设 $v \in OR$。那么，

$$Pr_c[N(v,c) \cap (R \setminus S) \neq \varnothing] \geq 1 - p.$$

证明（概要）。证明方法与推论 38 相同，只是我们使用引理 37 的第 2 部分，并利用 $\beta \geq \alpha(1-p)$，因为 $\beta \geq p$ 和 $\alpha \leq 1 - p$。 □

现在我们将根据必红顶点和强制蓝色顶点的数量分情况证明定理 32。

引理 40. 如果不存在必红顶点或不存在强制蓝色顶点，则该图可以连通。

证明。如果不存在必红顶点，我们就将每个顶点 $v \notin B \cup \{b\}$ 重新着色为颜色 c，使得 $N(v, c)$ 与 B 相交。因此，每个顶点都能直接或通过 B 中的一个顶点与 b 连通，由此证明该图是连通的。对称地，如果不存在强制蓝色顶点，则所有顶点都可以与 r 或 R 连通。 □

引理 41。 如果 $|OR| > C$ 或 $|OB| > C$，则该图可以是连通的。

证明。不失一般性，设 $|OB| > |OR|$（另一种情况是对称的）。令$OB' \subset OB \setminus \{b\}$ 为 C 中必为蓝色顶点的一个子集。我们将重新着色 OB' 中的所有顶点以及所有非必然顶点，使它们能够连接到 $B \setminus OB'$。我们将证明，OB' 中的每个顶点都存在一种颜色，使其既能连接到 $B \setminus OB'$，又能连接到 OR中剩余顶点的一个常数比例。由于 C 关于 n是对数级的，因此我们能够以这种方式覆盖 OR 中的所有顶点，从而使图连通。图 4 展示了这一过程的示意图。

对于任意 $u \in OB'$ 和 $v \in OR$，如果让它们各自从相应的分布中独立随机选取一种颜色，则有 $Pr[uv \in E(G)] \geq p' = p + \varepsilon$。此外，根据推论 38，$Pr[N(u,c) \cap (B \setminus OB') \neq \varnothing] \geq (1-p)$。因此，令 $Y_{u,v}$ 为指示随机变量，当且仅当 $uv \in E(G)$ 且 $N(u,c) \cap (B \setminus OB') \neq \varnothing$ 时取值为 1，我们有

$$Pr[Y_{u,v} = 1] \geq p' + (1-p) - 1 \geq \varepsilon.$$

集合 $Y_u = \sum_{v \in OR} Y_{u,\ v}$并注意，如果 $N(u,\ c)$ 与 $B \setminus OB'$ 不相交，那么 $Y_u = 0$；否则 Y_u 统计 u 在 OR 中的邻居（依据 OR 的新着色，而不是那个

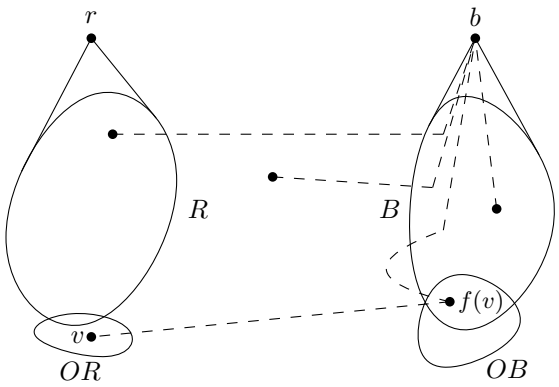


图5：如何根据从 OR 到 $OB \setminus \{b\}$的单射 f 连接图。

我们一开始的取值）。从各自的分布中独立随机选择 u 和所有 $v \in OR$ 的颜色，根据期望的线性性，我们得到

$$\mathbb{E}[Y_u] = \sum_{v \in OR} \mathbb{E}[Y_{u,v}] \geq \varepsilon |OR|.$$

这意味着，对于 u，存在一种颜色 c，并且可以为 OR 中的顶点选择颜色，使得 $Y_u \geq \varepsilon|OR|$；也就是说，使得 $N(u, c) \cap (B \setminus OB') \neq \varnothing$，并且根据 OR的新着色，u在 OR中至少有 $\varepsilon|OR|$ 个邻居。我们在 c 中为 u 着色，并根据新的着色固定分配给 u 的这些邻居的颜色。因此，OR 中的所有这些至少有 $\varepsilon|OR|$ 个顶点都成为 u的邻居，并且 u 与 $B \setminus OB'$中的某个顶点有一条边相连。

通过使用 OB'中的一个顶点，我们从而连接了 OR中占 ε比例的部分。我们重复这一过程，直到 OR 中不再有顶点为止。注意，我们始终可以至少匹配一个顶点；一旦 $\varepsilon|OR|$ 变得小于 1，就必须将这一点纳入考虑。因此，我们至多使用

$$1 + \varepsilon^{-1} + \log_{\frac{1}{1-\varepsilon}}(|OR|) \leq 1 + \varepsilon^{-1} + \frac{\log_2(n)}{\log_2(\frac{1}{1-\varepsilon})} \leq C$$

从 OB'中选取顶点，并相对于 ε将 $C'(\varepsilon)$取得足够大。

$|OR| > |OB|$ 的情形以对称的方式成立，使用推论 39。 □

因此，下面我们假设必红顶点和必为蓝色顶点的数量都大于 0 且至多为 C。

引理42。 如果存在从 OR 到 $OB \setminus \{b\}$的单射，或存在从OB 到 $OR \setminus \{r\}$的单射，则该图可以连通。

证明。 首先假设存在从 OR 到 $OB \setminus \{b\}$ 的单射 f。相反的情形以对称的方式成立。在这种情况下，我们旨在将每个顶点连接到 b，方法是将每个强制红色顶点 v 连接到顶点 $f(v)$。然后，将这些强制蓝色顶点以及 $B \setminus \{b\}$ 之外的非必然顶点连接到 $B \setminus OB$中的某个顶点。由于这些顶点与 b 相连，整个图便连通了。这一点如图5所示。注意，我们会重新为OB 中的顶点以及 B之外的顶点着色，但不会重新为 $B \setminus OB$中的顶点着色。

与引理 41 的证明相同，我们让 $v \in OR$ 和 $f(v) \in OB$ 独立地选择随机颜色。然后，根据推论 38，并且由于 $Pr[vf(v) \in E(G)] \geq p'$，使得 v 与 $f(v)$ 相连且 $f(v)$ 与 $B \setminus OB$ 相连的概率再次非零。因此，我们可以为每一对顶点（v，$f(v)$）选择这样的颜色，其中 $v \in OR$。

□

我们可以通过考虑非必然顶点来增强前面的引理：这些顶点也有较大的概率（至少为 $1-p$）稳健地连接到集合 R 和 B 中的任意一个（正如推论 38 和推论 39 所保证的，必然顶点分别能够做到这一点）：

引理43。 如果存在一个非强制顶点 $v \notin \{r,b\}$，使得 $Pr_c[|N(v、\ c) \cap S| > C] \geq 1-p$对于 $S = R$ 或 $S = B$成立，则该图可以连通。

证明。 如果满足引理 41 或引理42的条件，我们就可以使该图连通。否则，$|OR| = |OB| \leq C$、$r \in OR$ 和 $b \in OB$。假设存在一个顶点 v，使得$Pr_c[|N(v,\ c)\cap B| > C] \geq 1-p$（$S = R$的情况可对称地处理）。我们可以构造一个从 OR到 $(OB\cup\{v\})\backslash\{b\}$的双射。由于我们有 $Pr_c[N(v,\ c)\cap(B\backslash(OB\cup\{v\}\backslash\{b\})) \neq \varnothing] \geq 1-p$，引理42的证明同样适用于此映射。 □

此外，如果存在一个可用于连接红色集合和蓝色集合的顶点，我们也可以将图连接起来：

引理44。 如果存在一个顶点 $v \notin \{r,b\}$ 和一种颜色 c，使得 $N(v,c) \cap R \neq \varnothing$ 和$N(v,c) \cap B \neq \varnothing$，我们就可以连接该图。

证明。 我们给顶点 v赋予颜色 c。每个顶点 $w \notin R \cup B \cup \{r,b,v\}$ 都会被重新着色为某种颜色 c_w，使得 $|N(w,c_w)| > pn$。由于 $p > 1-\alpha-\beta$，w随后会与 R或 B中的某个顶点连通。对所有 w完成此操作后，该图就是连通的。 □

最后，我们将考虑前述引理未覆盖的情形；对于这些情形，我们需要以下关于非必然顶点的下界。

引理45。 对于任意非强制顶点 $v \notin \{r,b\}$，我们定义

$$p_v^r := \max_{c:|N(v,c)\cap B|\leq C} |N(v,c)|,$$
$$p_v^b := \max_{c:|N(v,c)\cap R|\leq C} |N(v,c)|.$$

假设不存在满足引理*43*或引理*44*中任一条件的顶点，则有$p_v^r > \frac{\beta p-\beta+p}{p} n$，且 $p_v^b > \frac{\alpha p-\alpha+p}{p} n$。

证明。 令 v 为任意一个非强制顶点，并令 $q_v := Pr_{c_v}[|N(v,c_v) \cap R| > C]$ 和 $z_v := Pr_{c_v}[|N(v,c_v) \cap B| > C]$。由于 v 不满足引理43的条件，因此有$q_v, z_v < 1-p$。现在我们考察选取一个随机颜色 c_v 时 $N(v,c_v)$ 的期望大小。首先，根据引理 34，有 $pn < \mathbb{E}_{c_v}[|N(v,c_v)|]$。为了得到 p_v^b 的界，我们还可以利用以下事实：

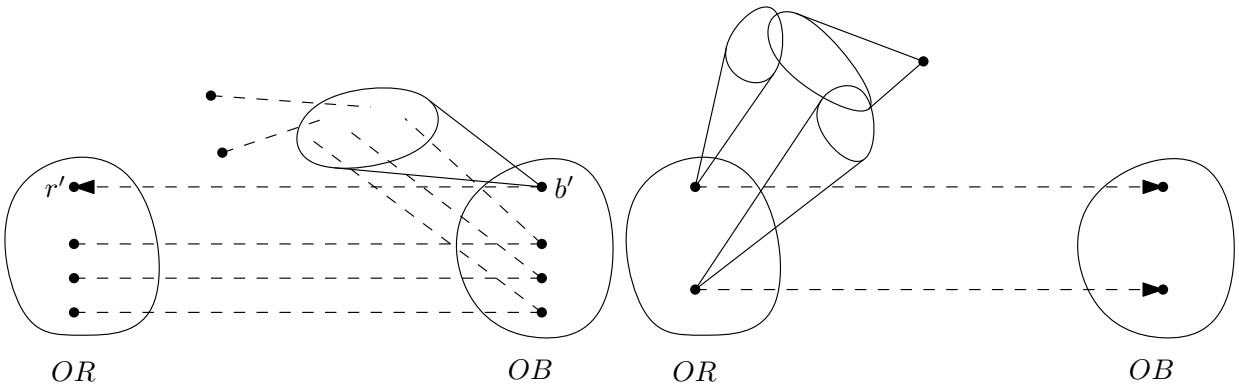
$$\mathbb{E}_{c_v}[|N(v,c_v)|] \leq q_v\alpha n + (1-q_v)p_v^b.$$

这是因为每个邻域的大小至多为 αn，并且根据 p_v^b的定义，v 中与 R 相交的顶点数至多为 C 的任意邻域，其大小至多为 p_v^b。结合这些界，我们得到$pn < q_v\alpha n + (1-q_v)p_v^b$，这意味着

$$p_v^b > \frac{p-q_v\alpha}{1-q_v} n = \left(\alpha - \frac{\alpha-p}{1-q_v}\right) n \geq \left(\alpha - \frac{\alpha-p}{p}\right) n = \frac{\alpha p-\alpha+p}{p} n,$$

如所愿。

为了得到关于 p_v^r 的界，我们可以使用界 $\mathbb{E}_{c_v}[|N(v,c_v)|] \leq z_v\beta n + (1-z_v)p_v^r$。这里，部分 $z_v\beta$ 是由以下事实得出的：任何与 B 相交的邻域都不可能与 R 相交（因为我们



（a）情形1：至少有一条边朝向 OR。（b）情形2：所有边都朝向 OB。

图6：连接引理46中的图。

（假设不存在引理44中所述的顶点），因此在选择顶点对 $(b,\ c_b)$ 时已将其考虑在内。因此，我们有 $pn < z_v\beta n + (1-z_v)p_v^r$，这意味着

$$p_v^r > \frac{p - z_v\beta}{1 - z_v} n = \left(\beta - \frac{\beta - p}{1 - z_v}\right) n \geq \left(\beta - \frac{\beta - p}{p}\right) n = \frac{\beta p - \beta + p}{p} n,$$

完成证明。 □

最后，我们覆盖唯一剩下的情形。

引理46。 如果有 $|OB| = |OR|$，且不存在引理*43*或引理*44*中所述的顶点，我们可以连接该图。

证明。 我们考虑 OR 与 OB之间的候选边。通过考察以下两种概率，可以将每条这样的边 $r'b'$ 定向：

$$p_{r'\to b'} := Pr_{c_{r'}}[\exists c_{b'} \text{ such that } r'b' \in E(G) \text{ in coloring } c(r') = c_{r'}, c(b') = c_{b'}]$$
$$p_{b'\to r'} := Pr_{c_{b'}}[\exists c_{r'} \text{ such that } r'b' \in E(G) \text{ in coloring } c(r') = c_{r'}, c(b') = c_{b'}]$$

注意，$p_{r'\to b'} \cdot p_{b'\to r'} \geq Pr[r'b' \in E(G)] \geq p'$，因此 $\max(p_{r'\to b'}, p_{b'\to r'}) \geq \sqrt{p'}$。现在，如果 $p_{r'\to \overline{b'}} > p_{b'\to r'}$，我们就将候选边 $r'b'$ 从 r' 定向到 b'；否则，从 b' 定向到 r'。我们使用 $|OR| = |OB|$，在 OR 与 OB之间的候选边中任意选取一个完美匹配 M。

情形 1： 存在一条弧 $(b', r') \in M$，其方向指向 $r' \in OR$。在这种情况下，我们将按图6a所示连接该图。每个强制红色顶点都与其匹配的必蓝顶点相连。顶点 b' 必须具有较大的邻域大小，这一点以概率 $\geq 1 - \sqrt{p}$得到保证，从而使其能够同时与 r'相连。每个必蓝顶点和每个非强制顶点都与该邻域相连。

由引理 37 的第 3 部分可得

$$Pr[|N(b', c)| \geq (1-p)(1-\beta)n + \varepsilon n/2] \geq 1 - \sqrt{p}.$$

因此，对于 b' 的颜色选择，以正概率，b' 的邻域大小至少为 $(1-p)(1-\beta)n + \varepsilon n/2$，并且存在一种 $\underline{r'}$ 的颜色选择，使得 $r'b'$ 是一条边，因为 $1 - \sqrt{p} + \sqrt{p'} > 0$。

同样，对于所有其他必为蓝色的顶点 u，根据引理 37 的第 1 部分，我们有

$$Pr[|N(u,c)| \geq \frac{p+p\beta-\beta}{p}n + \varepsilon n/2] \geq 1-p,$$

因此，以正概率存在一条边连接 u 与其在 OR 中的匹配顶点，并且 u的邻域大小至少为 $\frac{p+p\beta-\beta}{p}n + \varepsilon n/2$。

我们需要证明，这些有保证大小的邻域必须在 $OB \cup OR$之外相交。令 $S := (1-p)(1-\beta)n + \varepsilon n/2$ 以及 $T := \frac{p+p\beta-\beta}{p}n + \varepsilon n/2$。由于两个邻域与 R 相交的顶点数最多为 ~~C，~~ 我们需要证明$(S-3C) + (T-3C) > n - |R|$，即

$$S + T + \alpha n - n - 6C \overset{!}{>} 0. \tag{3}$$

此外，为了保证 b' 邻域之外的每个非强制顶点 u 都能连接到该邻域，我们需要保证 u 的邻域与 b' 的邻域在 $OB \cup OR$之外相交。为此，我们需要检验$(S-3C) + (p_u^b - 3C) > n - |R|$，即

$$S + p_u^b + \alpha n - n - 6C \overset{!}{>} 0. \tag{4}$$

由于我们对 p_u^b 的下界小于 T（因为 $\alpha \geq \beta$），(4) 推出 (3)，因此只需证明 (4)。

$$\begin{aligned}
S + p_u^b + \alpha n - n - 6C &\geq (1-\beta)(1-p)n + \frac{p+\alpha p-\alpha}{p}n + \alpha n - 6C - n + \varepsilon n/2 \\
&= \left(-\beta - p + \beta p + 1 + 2\alpha - \frac{\alpha}{p} - 6\frac{C}{n} + \varepsilon/2\right) n \\
&= \left(1 - p - \alpha(1/p - 2) - \beta(1-p) - 6\frac{C}{n} + \varepsilon/2\right) n \\
&= \left(1 - p - (\alpha+\beta)(1-p) - 6\frac{C}{n} + \varepsilon/2\right) n \\
&\geq \varepsilon n/2 - 6C,
\end{aligned}$$

当 n 足够大时，该值大于 0。注意，这里我们用到了观察 33。

情形2： M 中的所有边都朝向 OB。在这种情况下，我们将按图6b所示使该图连通。我们为每个强制红色顶点赋予一个较大的邻域，从而仍能保证强制蓝色顶点可以与其匹配伙伴相连。随后，我们为每个尚未与任何强制顶点直接相连的非强制顶点赋予一个较大的邻域，且该邻域不与 B 相交（由引理45保证）。由于这些邻域的规模足够大，它们必须与所有强制红色顶点的邻域相交，因此该图是连通的。

令 $Y := (1-p)(1-\alpha)n + \varepsilon n/2$。对于任意强制红色顶点 v，由引理 37 的第 4 部分可得$Pr[|N(v,\ c_v)| \geq Y] \geq 1 - \sqrt{p}$，因此，以概率$\neg 1 - \sqrt{p} + \sqrt{p'} > 0$，存在一种颜色可赋给 v 的强制蓝色伙伴顶点，使二者由一条边相连，并且还有 $|N(v,\ c_v)| \geq Y$。

现在我们检查，每个非强制顶点 u 的邻域都与每个大小至少为 Y 的必红顶点邻域相交。回想一下，根据引理45，每个 u都有一个无法与大小为 p_u^r的 B 稳健相交的邻域。注意，我们需要证明这些邻域在 $OR \cup OB$ 之外相交，因为这些顶点可能会改变颜色。为此，只需证明 $(p_u^r - 3C) + (Y - 3C) > n - |B|$，即

$$p_u^r + Y + \beta n - n - 6C \overset{!}{>} 0.$$

我们计算

$$\begin{aligned} p_u^r + Y + \beta n - n - 6C &\geq \left(\frac{\beta p - \beta + p}{p} + (1-\alpha)(1-p) + \beta - 1 - 6\frac{C}{n} + \varepsilon/2 \right) n \\ &= \left(\beta - \frac{\beta}{p} + 1 - \alpha - p + \alpha p + \beta - 6\frac{C}{n} + \varepsilon/2 \right) n \\ &= \left(1 - p + \beta(2 - 1/p) - \alpha(1-p) - 6\frac{C}{n} + \varepsilon/2 \right) n \\ &= \left(1 - p - (\alpha+\beta)(1-p) - 6\frac{C}{n} + \varepsilon/2 \right) n \\ &\geq \varepsilon n/2 - 6C, \end{aligned}$$

当 n 足够大时，该值大于 0。 □

我们已经涵盖了所有可能的情况，因此现在可以证明定理 32。

定理 32 的证明。 引理40至44和引理46涵盖了所有可能的情况。因此，我们总能使图连通，从而证明该定理。 □

定理 7 的证明。 定理 7 可由引理20、21、25、26及28至31、推论22和24，以及定理 32 推出。 □

参考文献

[1] Jon Aaronson、David Gilat、Michael Keane 和 Vincent de Valk。《一类一依赖过程的代数构造》。概率年刊，第128–143页，1989年。[2] Noga Alon 和 Asaf Nussboim。k-重独立随机图。载于《2008年第49届IEEE计算机科学基础年会论文集》，第813–822页，2008年。doi:10.1109/ FOCS.2008.61。[3] Noga Alon 和 Joel H. Spencer。《概率方法》。John Wiley & Sons，2016年。doi: 10.1002/9780470277331。[4] Kazuoki Azuma。某些相关随机变量的加权和。《东北数学杂志》，19(3):357–367，1967年。doi:10.2748/tmj/1178243286。[5] Leila Badakhshian、Victor Falgas-Ravry 和 Maryam Sharifzadeh。多部图中横截树的密度条件。arXiv预印本，2023年。doi:10.48550/arXiv.2305. 05713。[6] Paul Balister、B´ela Bollob´as 和 Mark Walters。随机收发网络。《应用概率进展》，41(2):323–343，2009年。doi:10.1239/aap/1246886613。[7]Paul Balister、Tom Johnston、Michael Savery 和 Alex Scott。关于 $\mathbb{Z}^n$上的1-独立渗流的改进界。arXiv预印本，2022年。doi:10.48550/arXiv.2206.12335。[8] B´ela Bollob´as。随机正则图的等周数。《欧洲组合学杂志》，9(3):241–244，1988年。doi:10.1016/S0195-6698(88)80014-3。[9] Adrian Bondy、Jian Shen、St´ephan Thomass´e 和 Carsten Thomassen。多部图中三角形的密度条件。《组合学》，26(2):121–131，2006年。doi:10.1007/ s00493-006-0009-y。[10] Karl Bringmann、Ralph Keusch 和 Johannes Lengler。几何非齐次随机图。《理论计算机科学》，760:35–54，2019年。doi:10.1016/j.tcs.2018.08.014。[11] Robert M. Burton、Marc Goulet 和 Ronald Meester。关于一依赖过程和 k-块因子。《概率年刊》，21(4):2157–2168，1993年。doi:10.1214/aop/1176989014。.[12] Constantin Carath´eodory。关于不取给定值的幂级数系数的变动范围。《数学年刊》，64(1):95–115，1907年。doi:10. 1007/BF01449883。[13] Fan Chung 和 Linyuan Lu。具有给定期望度数序列的随机图中的连通分量。《组合学年刊》，6(2):125–145，2002年。doi:10.1007/PL00012580。[14] P´eter Csikv´ari 和 Zolt´an L. Nagy。密度 Tur´an 问题。《组合学、概率与计算》，21(4):531–553，2012年。doi:10.1017/S0963548312000016。[15] A. Nicholas Day、Victor Falgas-Ravry 和 Robert Hancock。1-独立随机图中的长路径与连通性。《随机结构与算法》，57(4):1007–1049，2020年。doi:10.1002/rsa.20972。[16] Maria Deijfen、Remco van der Hofstad 和 Gerard Hooghiemstra。无标度渗流。《IHP概率与统计年刊》，49(3):817–838，2013年。doi:10.1214/12-AIHP480。

[17] 保罗·埃尔德什。图论与概率。《加拿大数学杂志》，11:34–38，1959年。doi:10.4153/CJM-1959-003-9。[18] Victor Falgas-Ravry 和 Vincent Pfenninger。$\mathbb{Z}^2 \times K_n$上的1-独立渗流。《随机结构与算法》，2022年。doi:10.1002/rsa.21129。[19] Heidi Gebauer、Robin A. Moser、Dominik Scheder 和 Emo Welzl。Lov´asz 局部引理与可满足性。载于 Susanne Albers、Helmut Alt 和 Stefan N¨aher（编），《高效算法：献给 Kurt Mehlhorn 六十岁生日的论文集》，第30–54页。Springer，柏林、海德堡，2009年。doi:10.1007/978-3-642-03456-5_3。[20] Alexander E. Holroyd 和 Thomas M. Liggett。有限依赖着色。载于《数学论坛：Pi》，第4卷，第e9页。剑桥大学出版社，2016年。doi:10.1017/fmp. 2016.7。[21] Svante Janson、Andrzej Ruci´nski 和 Tomasz ´Luczak。《随机图》。Wiley 离散数学与优化丛书。Wiley，2011年。doi:10.1002/9781118032718。[22] Anders Johansson。无三角形图的渐近选择数。技术报告，DIMACS，第91–95页，1996年。[23] Brian Karrer 和 Mark E. J. Newman。网络中的随机区块模型与社区结构。《物理评论 E》，83(1):016107，2011年。doi:10.1103/PhysRevE.83.016107。[24] Dmitri Krioukov、Fragkiskos Papadopoulos、Maksim Kitsak、Amin Vahdat 和 Mari´an Bo- gun´a。复杂网络的双曲几何。《物理评论 E》，82(3):036106，2010年。doi:10.1103/PhysRevE.82.036106。[25] Thomas M. Liggett、Roberto H. Schonmann 和 Alan M. Stacey。乘积测度的支配。《概率年刊》，25(1):71–95，1997年。doi:10.1214/aop/1024404279。[26] Michael Molloy。团数较小的图的列表色数。《组合理论杂志 B辑》，134(1):264–284，2019年。doi:10.1016/j.jctb.2018.06.007。[27] Zolt´an L. Nagy。Tur´an 问题的一个多部图版本——密度条件与特征值。《组合学电子杂志》，18(1):P46，2011年。doi:10.37236/533。[28] Mathew Penrose。《随机几何图》。牛津大学出版社，2003年。doi:10.1093/acprof:oso/9780198506263.001.0001。[29] Hannu Reittu 和 Ilkka Norros。通信网络拓扑的随机图模型。载于《复杂系统的统一主题 VII：第七届国际复杂系统会议论文集》，第214–221页。Springer，2012年。doi:10.1007/978-3-642-18003-3_21。[30] James B. Shearer。关于 Spencer 的一个问题。《组合数学》，5(3):241–245，1985年。doi: 10.1007/BF02579368。[31] Joel Spencer。Ramsey 函数的渐近下界。《离散数学》，20:69–76，1977年。doi:10.1016/0012-365X(77)90044-9。

A $G_{n,\ 1/2}$ 带有附加约束

定义 47。 对于整数 k，令 $G^k_{n,\ 1/2}$ 为通过将 $G_{n,\ 1/2}$ 限定在所得图中边的数量可被 k 整除这一事件上而得到的关于含 n 个顶点的图的概率分布。

注意，由于 $G_{n,\ 1/2}$ 是在所有具有 n 个顶点的图上的均匀分布，因此 $G^k_{n,\ 1/2}$ 是在所有具有 n 个顶点且边数能被 k 整除的图上的均匀分布。我们将证明，$G^2_{n,\ 1/2}$ 是一个着色模型，而 $G^3_{n,\ 1/2}$ 甚至不是两两独立的。

引理 48. 对于任意 $n \in \mathbb{N}$，分布 $G^2_{n,\ 1/2}$ 是一个着色模型。

证明。 我们通过给出一种着色模型，并证明它等价于$G^2_{n,\ 1/2}$，来证明这个结论。对于每个顶点 v，在所有含有偶数个 1 的字符串中均匀随机地抽取一个字符串 $s_v \in \{0,1\}^{n-1}$。字符串 s_v 是 v的颜色。 s_v 中的每一位都对应一个不同的其他顶点。对于顶点 $u \neq v$，我们将 $s_v(u)$ 记作 s_v 中对应于顶点 u的那一位。当且仅当 $s_v(u) + s_u(v) = 1$ 时，边 $\{u,v\}$ 才会出现在我们的图中。

我们首先证明，以这种方式采样得到的图始终具有偶数条边。注意，当且仅当 $s_v(u) + s_u(v) \equiv 1 \bmod 2$ 时， $\{u,v\}$ 才存在。因此，边的总数的奇偶性为

$$\sum_{u \neq v} \big(s_v(u) + s_u(v)\big) \mod 2 = \sum_v \sum_{u \neq v} s_v(u) \mod 2 = \sum_u 0 \mod 2 = 0 \mod 2,$$

按要求。

为了证明每个具有偶数条边的图都是等概率的，注意到每一种有效的顶点着色都是等概率的，并且每个具有偶数条边的图都恰好可以由 precisely$2^{(n-2)(n-1)/2}$ 种不同的顶点着色产生。对于后者，考虑某个具有偶数条边的图 G，其顶点为 $v_1, \ldots, v_n$。回顾一下， $X_{\{u,v\}}$是指示变量，表示 $\{u,v\} \in E(G)$。注意，一般而言，对于任意两个顶点 $u \neq v$，$s_u(v)$取0或1都是可能的，但每一种这样的选择都会通过 $s_v(u) = X_{\{u,v\}} - s_u(v) \bmod 2$ 唯一确定 $s_v(u)$。因此， s_{v_1}有 2^{n-2} 种有效选择，每一种都与 G一致，因为在 $n-1$ 位字符串中，恰好有 2^{n-2}个字符串包含偶数个1。固定 v_1的颜色为 s_{v_1}，如上所述，这就唯一确定了每个 $i \in \{2,\ldots,n\}$对应的 $s_{v_i}(v_1)$。因此，s_{v_2}有 2^{n-3} 种有效选择，每一种都会确定每个 $i \geq 3$对应的 $s_{v_i}(v_2)$。以相同的方式继续，对于每个 $i \in [n-1]$， s_{v_i} 都有 2^{n-i-1} 种有效选择，而 s_{v_n}只有一种有效选择。总而言之，能够生成 G 的着色数量为

$$\prod_{i=1}^{n-2} 2^i = 2^{\sum_{i=1}^{n-2} i} = 2^{(n-2)(n-1)/2},$$

这不依赖于 G，符合要求。 □

Le引理 49。 设 $n \geq 4$。那么分布 $G^3_{n,\ 1/2}$ 并非两两独立。

证明。 考虑两条顶点不相交的边 e 和 f。只需证明

$$Pr[X_e | X_f] \neq Pr[X_e | \overline{X_f}],$$

因为这将意味着

$$Pr[X_e] = Pr[X_e|X_f]Pr[X_f] + Pr[X_e|\overline{X_f}]Pr[\overline{X_f}] \neq Pr[X_e|X_f].$$

令$G \sim G^3_{n,\ 1/2}$，并令 $s \in \{0,1\}^{\binom{n}{2}}$ 描述 G 中的边。具体而言，在 G 的所有 $m = \binom{n}{2}$ 候选边上选取某种顺序e_1, ..., e_m，使得 $e_1 = f$ 且 $e_2 = e$，并令 $s_i = 1$ 当且仅当$e_i \in E(G)$。用 $w(s)$ 表示字符串 s 中 1 位的数量，对于任意 $r \in \{0,1,2\}$ 和 k，令

$$S^r_k = \{s \in \{0,1\}^k | w(s) \equiv r \mod 3\}.$$

然后，对$G \sim G^3_{n,\ 1/2}$进行采样，等价于从 S^0_m 中均匀随机地选取 s。我们有

$$Pr[X_e|X_f] = \frac{Pr[X_e \text{ and } X_f]}{Pr[X_f]} = \frac{|\{s \in S^0_m | s_1 = 1 \text{ and } s_2 = 1\}|}{|\{s \in S^0_m | s_1 = 1\}|}$$

$$= \frac{|\{s \in S^0_m | s_1 = 1 \text{ and } s_2 = 1\}|}{|\{s \in S^0_m | s_1 = 1 \text{ and } s_2 = 1\}| + |\{s \in S^0_m | s_1 = 1 \text{ and } s_2 = 0\}|} = \frac{|S^1_{m-2}|}{|S^1_{m-2}| + |S^2_{m-2}|}$$

and

$$Pr[X_e|\overline{X_f}] = \frac{Pr[X_e \text{ and } \overline{X_f}]}{Pr[\overline{X_f}]} = \frac{|\{s \in S^0_m | s_1 = 0 \text{ and } s_2 = 1\}|}{|\{s \in S^0_m | s_1 = 0\}|}$$

$$= \frac{|\{s \in S^0_m | s_1 = 0 \text{ and } s_2 = 1\}|}{|\{s \in S^0_m | s_1 = 0 \text{ and } s_2 = 1\}| + |\{s \in S^0_m | s_1 = 0 \text{ and } s_2 = 0\}|} = \frac{|S^2_{m-2}|}{|S^2_{m-2}| + |S^0_{m-2}|}.$$

因此，只需证明♀♀♀♀

$$\frac{|S^1_{m-2}|}{|S^1_{m-2}| + |S^2_{m-2}|} \overset{!}{\neq} \frac{|S^2_{m-2}|}{|S^2_{m-2}| + |S^0_{m-2}|}$$

$$|S^1_{m-2}||S^2_{m-2}| + |S^1_{m-2}||S^0_{m-2}| \overset{!}{\neq} |S^2_{m-2}||S^1_{m-2}| + |S^2_{m-2}||S^2_{m-2}|$$

$$|S^1_{m-2}||S^0_{m-2}| \overset{!}{\neq} |S^2_{m-2}||S^2_{m-2}|. \tag{5}$$

注意，$|S^0_{m-2}| + |S^1_{m-2}| + |S^2_{m-2}| = 2^{m-2} \neq 0$ 模 3，这意味着不可能有$|S^0_{m-2}| = |S^1_{m-2}| = |S^2_{m-2}|$。

如果 n 或 $n-1$ 能被 3 整除，那么 $m = \binom{n}{2}$也能被 3 整除，因此 $m - 2 \equiv 1$ 模 3。剩下的情况是 $n \equiv 2$ 模 3，这意味着 $m = \binom{n}{2} \equiv 1$ 模 3，因此 $m - 2 \equiv 2$ 模 3。

如果 $m - 2 \equiv 1 \mod 3$，则存在一个双射 $f\colon S^0_{m-2} \to S^1_{m-2}$，它是通过翻转所有比特得到的；也就是说，对于 $s \in S^0_{m-2}$，对每个 $i \in [m-2]$，都有 $\big(f(s)\big)_i = 1 - s_i$。因此 $|S^0_{m-2}| = |S^1_{m-2}|$，所以只要 $|S^1_{m-2}| \neq |S^2_{m-2}|$，等式 (5) 就成立。我们知道这一条件成立，因为并非所有 $|S^r_{m-2}|$ 都相同。

类似地，如果 $m - 2 \equiv 2 \mod 3$，则通过翻转所有比特，可以得到一个从 S^0_{m-2} 到 S^2_{m-2} 的双射。因此，$|S^0_{m-2}| = |S^2_{m-2}|$，并且只要 $|S^1_{m-2}| \neq |S^2_{m-2}|$，等式 (5) 仍然成立，而这同样必然成立。

□

B 引理 13 的证明

引理 *13* 的证明。令 $F := \alpha \cdot C + (1-\alpha) \cdot D$ 为我们所考虑的 C 与 D 的凸组合。首先注意，对于任意边 e，我们有 $Pr_F[X_e] = \alpha Pr_C[X_e] + (1-\alpha) Pr_D[X_e] = Pr_C[X_e] = Pr_D[X_e]$。

首先假设 C 和 D 两两独立。考虑一对不相邻的边e, f。我们有

$$Pr_F[X_e \text{ and } X_f] = \alpha Pr_C[X_e \text{ and } X_f] + (1-\alpha) Pr_D[X_e \text{ and } X_f]$$

$$= \alpha Pr_C[X_e] Pr_C[X_f] + (1-\alpha) Pr_D[X_e] Pr_D[X_f] = Pr_F[X_e] Pr_F[X_f],$$

其中第二个等式由 C 和 D的两两独立性得出，第三个等式则由以下事实得出：对于每条边 e，$Pr_F[X_e] = Pr_D[X_e] = Pr_C[X_e]$，如上所述。这表明 F 也具有两两独立性。

现在假设 C 和 D 均为匹配独立。要证明 F 也匹配独立，只需证明：对于任意由顶点不相交的边 $e_1, \ldots, e_k$组成的匹配 M 以及任意子集 $S \subseteq M$，都有 $Pr_F[X_S$和 $\overline{Y_{M\setminus S}}] = \prod_{e\in S} Pr_F[X_e] \prod_{e\in M\setminus S} Pr_F[X_e]$，其中 X_S 是 S 中的所有边均存在这一事件，而 $Y_{M\setminus S}$ 是 $M \setminus S$ 中的边均不存在这一事件。我们有

$$Pr_F[X_S \text{ and } Y_{M\setminus S}] = \alpha Pr_C[X_S \text{ and } Y_{M\setminus S}] + (1-\alpha) Pr_D[X_S \text{ and } Y_{M\setminus S}]$$

$$= \alpha \prod_{e\in S} Pr_C[X_e] \prod_{e\in M\setminus S} Pr_C[\overline{X_e}] + (1-\alpha) \prod_{e\in S} Pr_D[X_e] \prod_{e\in M\setminus S} Pr_D[\overline{X_e}]$$

$$\prod_{e\in S} Pr_F[X_e] \prod_{e\in M\setminus S} Pr_F[\overline{X_e}],$$

这里我们再次利用了 C 和 D 的匹配独立性，以及 $Pr_F[X_e] = Pr_D[X_e] = Pr_C[X_e]$这一事实。

最后，假设 C 和 D 满足边子图独立性。设 $e = \{u, v\}$ 为一条边，W为一个顶点集，且 u, $v \notin W$，再设 H_W为 W上的诱导子图的一种选择。我们将证明 F 也满足边子图独立性，即

$$Pr_F[G[W] = H_W \text{ and } X_e] = Pr_F[G[W] = H_W] Pr_F[X_e].$$

我们有

$$Pr_F[G[W] = H_W \text{ and } X_e] = \alpha Pr_C[G[W] = H_W \text{ and } X_e] + (1-\alpha) Pr_D[G[W] = H_W \text{ and } X_e]$$

$$= \alpha Pr_C[G[W] = H_W] Pr_C[X_e] + (1-\alpha) Pr_D[G[W] = H_W] Pr_D[X_e]$$

$$= \Big(\alpha Pr_C[G[W] = H_W] + (1-\alpha) Pr_D[G[W] = H_W]\Big) Pr_F[X_e] = Pr_F[G[W] = H_W] Pr_F[X_e],$$

如所需，这里用到了 C 和 D的边子图独立性。 □

C 断言 19 的证明

断言 *19* 的证明。 我们固定一条候选边 $e = \{u, v\}$。于是有

$$p = Pr[X_e] = q\frac{n-2}{n} + (1-q)\frac{1}{n-1},$$

因为在团情形中，只有当 u 或 v 被选为孤立顶点 x时， e 才不会出现；而在匹配情形中，顶点 u 有 $n-1$ 种可选顶点与之匹配，每种选择出现的

概率相同。现在考虑两条可能的顶点不相交的边 $e=\{u,v\}$ 和 $f=\{w,z\}$。类似地，我们有

$$Pr[X_e \text{ and } X_f] = q\frac{n-4}{n} + (1-q)\frac{1}{n-1}\frac{1}{n-3},$$

其中最后一项来自这样一个事实：在匹配情形下，已知 u 和 v 匹配时，w 和 z 匹配的概率为 $\frac{1}{n-3}$。

为了实现两两独立性，只需找到 $q(n)$，使得 $Pr[X_e \text{和 } X_f] = Pr[X_e]Pr[X_f]$，即

$$\Big(q\frac{n-2}{n} + (1-q)\frac{1}{n-1}\Big)^2 \stackrel{!}{=} q\frac{n-4}{n} + (1-q)\frac{1}{n-1}\frac{1}{n-3}$$

$$\left(q\Big(\frac{n-2}{n} - \frac{1}{n-1}\Big) + \frac{1}{n-1}\right)^2 \stackrel{!}{=} q\Big(\frac{n-4}{n} - \frac{1}{n-1}\frac{1}{n-3}\Big) + \frac{1}{n-1}\frac{1}{n-3}$$

$$q^2\Big(\frac{n-2}{n} - \frac{1}{n-1}\Big)^2 + 2q\Big(\frac{n-2}{n} - \frac{1}{n-1}\Big)\frac{1}{n-1} + \frac{1}{(n-1)^2} \stackrel{!}{=} q\Big(\frac{n-4}{n} - \frac{1}{n-1}\frac{1}{n-3}\Big) + \frac{1}{n-1}\frac{1}{n-3}$$

$$q^2\Big(\frac{n^2-4n+2}{n(n-1)}\Big)^2 + q\frac{-n^4+11n^3-40n^2+58n-24}{n(n-1)^2(n-3)} - \frac{2}{(n-1)^2(n-3)} \stackrel{!}{=} 0$$

$$q^2(n^5 - 11n^4 + 44n^3 - 76n^2 + 52n - 12) + q(-n^5 + 11n^4 - 40n^3 + 58n^2 - 24n) - 2n^2 \stackrel{!}{=} 0.$$

解关于 q 的二次方程，其中

$$\begin{aligned} a &= n^5 - 11n^4 + 44n^3 - 76n^2 + 52n - 12 \\ b &= -n^5 + 11n^4 - 40n^3 + 58n^2 - 24n \\ c &= -2n^2, \end{aligned}$$

得到

$$\begin{aligned} q_\pm &= \frac{-b \pm \sqrt{b^2-4ac}}{2a} \\ &= \frac{n^5 - 11n^4 + 40n^3 - 58n^2 + 24n \pm \sqrt{n^{10} - 22n^9 + 201n^8 - 988n^7 + \Theta(n^6)}}{2n^5 - 22n^4 + 88n^3 - 152n^2 + 104n - 24}. \end{aligned}$$

考虑

$$\begin{aligned} q_+ &= \frac{n^5 - 11n^4 + 40n^3 - 58n^2 + 24n + \sqrt{n^{10} - 22n^9 + 201n^8 - 988n^7 + \Theta(n^6)}}{2n^5 - 22n^4 + 88n^3 - 152n^2 + 104n - 24} \\ &= \frac{1}{2} - \frac{2}{n^2} + \Theta\Big(\frac{1}{n^3}\Big) + \frac{\sqrt{(n^5 - 11n^4 + 40n^3)^2 - 108n^7 + \Theta(n^6)}}{2n^5 - 22n^4 + 88n^3 - 152n^2 + 104n - 24} \\ &\le \frac{1}{2} - \frac{2}{n^2} + \Theta\Big(\frac{1}{n^3}\Big) + \frac{1}{2} - \frac{2}{n^2} + \Theta\Big(\frac{1}{n^3}\Big) = 1 - \frac{4}{n^2} + \Theta\Big(\frac{1}{n^3}\Big). \end{aligned}$$

还要注意，

$$q_+ = \frac{n^5 - 11n^4 + 40n^3 - 58n^2 + 24n + \sqrt{n^{10} - 22n^9 + 201n^8 - 988n^7 + \Theta(n^6)}}{2n^5 - 22n^4 + 88n^3 - 152n^2 + 104n - 24}$$
$$= \frac{1}{2} - \frac{2}{n^2} + \Theta\Big(\frac{1}{n^3}\Big) + \frac{\sqrt{(n^5 - 11n^4 + 40n^3 - 200n^2)^2 + 292n^7 + \Theta(n^6)}}{2n^5 - 22n^4 + 88n^3 - 152n^2 + 104n - 24}$$
$$\geq \frac{1}{2} - \frac{2}{n^2} + \Theta\Big(\frac{1}{n^3}\Big) + \frac{1}{2} - \frac{2}{n^2} + \Theta\Big(\frac{1}{n^3}\Big) = 1 - \frac{4}{n^2} + \Theta\Big(\frac{1}{n^3}\Big).$$

这表明，对于某个 $q$with $q = 1 - \frac{4}{n^2} + \Theta\Big(\frac{1}{n^3}\Big)$，分布 $CM(n,\ q)$ 是定义良好的，并且两两独立。于是

$$p = Pr[X_e] = q\frac{n-2}{n} + (1-q)\frac{1}{n-1}$$
$$= 1 - \Theta\Big(\frac{1}{n^2}\Big) - \frac{2}{n} + \Theta\Big(\frac{1}{n^3}\Big) = 1 - \frac{2}{n} - \Theta\Big(\frac{1}{n^2}\Big),$$

如所要求的。

最后，我们证明，对于该 q，$CM(n,\ q)$ 不是匹配独立的。为此，考虑由三条边 $e,\ f,\ g$ 构成的匹配。我们有

$$Pr[X_e \text{ and } X_f \text{ and } X_g] = q\frac{n-6}{n} + (1-q)\frac{1}{n-1}\frac{1}{n-3}\frac{1}{n-5}$$
$$= 1 - \frac{6}{n} - \frac{4}{n^2} + \Theta\Big(\frac{1}{n^3}\Big).$$

为了满足匹配独立性，该表达式必须等于

$$Pr[X_e]Pr[X_f]Pr[X_g] = \left(q\frac{n-2}{n} + (1-q)\frac{1}{n-1}\right)^3$$
$$= \Big(1 - \frac{2}{n}\Big)^3\Big(1 - \frac{4}{n^2}\Big)^3 + \Theta\Big(\frac{1}{n^3}\Big) = \Big(1 - \frac{6}{n} + \frac{12}{n^2}\Big)\Big(1 - \frac{12}{n^2}\Big) + \Theta\Big(\frac{1}{n^3}\Big)$$
$$= 1 - \frac{6}{n} + \Theta\Big(\frac{1}{n^3}\Big) > 1 - \frac{6}{n} - \frac{4}{n^2} + \Theta\Big(\frac{1}{n^3}\Big) = Pr[X_e \text{ and } X_f \text{ and } X_g],$$

从而表明，对于 q 的这一选择，$CM(n,\ q)$ 不是匹配独立的 . □

D 引理 37 的证明

引理 *37* 的证明。 对于该引理的第 1 部分，我们确定能够以至少 $1-p$ 的概率保证出现的每个必蓝顶点 u 的最大邻域大小 T。邻域大小的上界为 βn，期望大小至少为 $(p+\varepsilon/2)n$。因此，根据引理 10，对于满足以下条件的每个 T

$$\frac{pn + \varepsilon n/2 - T}{\beta n - T} \geq 1 - p \tag{6}$$

我们得到

$$Pr[|N(u, c_u)| > T] \geq 1 - p.$$

我们求解方程（6）中的 T:

$$T \leq \frac{pn + \varepsilon n/2 - \beta n(1-p)}{p}$$
$$T \leq \frac{p - \beta + \beta p}{p} n + \frac{\varepsilon n/2}{p}$$

特别地，因此我们可以保证存在一个大小为 $\frac{p-\beta+\beta p}{p} n + \varepsilon n/2 < \frac{p-\beta+\beta p}{p} n + \frac{\varepsilon n}{2p}$ 的邻域，其概率至少为 $1 - p$。我们可以仿照引理的第 1 部分证明第 2 部分，只需将 β替换为 α。

对于引理的第 3 部分，我们再次使用引理 10 来确定能够以至少 $1 - \sqrt{p}$的概率保证出现的 u的最大邻域大小 S。我们可以使用 βn 的上界，因为 u 必为蓝色，且期望至少为 $(p + \varepsilon/2)n$。因此，我们再次得到

$$Pr[|N(u, c_u)| > S] \geq 1 - \sqrt{p}$$

对于每个满足以下条件的 S

$$\frac{pn + \varepsilon n/2 - S}{\beta n - S} \geq 1 - \sqrt{p},$$

或者等价地

$$S \leq \frac{pn + \varepsilon n/2 - \beta n(1 - \sqrt{p})}{\sqrt{p}}$$
$$S \leq \frac{p(1-\beta)n + \varepsilon n/2}{1-p} = (1-p)(1-\beta)n + \frac{\varepsilon n/2}{1-p},$$

其中我们利用了观察 33 得到 $\sqrt{p} = 1 - p$。特别地，因此我们可以保证存在一个大小为 $(1-p)(1-\beta)n + \varepsilon n/2 < (1-p)(1-\beta)n + \frac{\varepsilon n/2}{1-p}$ 的邻域，其概率至少为 $1 - \sqrt{p}$。最后，第 4 部分可仿照第 3 部分证明，只需将 β 替换为 α。 □

E 着色模式只需有限多种颜色 l

引理 8 的证明。设 $\mathcal{D}$ 是一个以着色模型给出的随机图分布。我们将逐个减少每个顶点的颜色数，同时保持概率分布 $\mathcal{D}$ 不变。稍微滥用一下记号，这里我们将某个顶点处实验的任何结果都称为它的颜色，即使这样的结果有无穷多个。假设我们正在考虑顶点 v；对于每个具有 n 个顶点的图 H，以及顶点 v 的颜色 c，令 $p_{H,c}$ 表示在顶点 v 的颜色为 c 的条件下，从 $\mathcal{D}$ 中抽样得到 H 的概率，并令 p_H 表示从 $\mathcal{D}$ 中抽样得到 H 的无条件概率。现在，对于顶点 v 的每种颜色 c，考虑向量 $\vec{p}_{*,c}$，它们位于 $2^{\binom{n}{2}}$ 维空间中，其分量是在某个规范顺序下，所有具有 n 个顶点的可能图 H 对应的所有 $p_{H,c}$。再考虑一个维数相同的向量 $\vec{p}_*$，其分量是在同一规范顺序下所有具有 n 个顶点的可能图 H 对应的 p_H。根据全概率公式，$\vec{p}_*$ 属于所有这些 $\vec{p}_{*,c}$ 向量的凸包（这样的向量可能有无穷多个）。由 Carathéodory 定理 [12] 可知，$\vec{p}_*$ 是某些向量的凸组合；这些向量从 $\{\vec{p}_{*,c} \mid c$ 中选出，共有 $2^{\binom{n}{2}}$ 个，而 $+1$ 是顶点 $v\}$ 的一种颜色。因此，我们可以为 v 选取相应的 $2^{\binom{n}{2}}$ 种颜色 $+1$，将该凸组合给出的系数用作顶点 v 处实验中这些颜色的概率，最终得到相同的分布 $\mathcal{D}$。对每个顶点重复这一过程，即完成证明。

□