arXiv:2601.20020v1 [math.ST] 27 Jan 2026

匹配与混合：马尔可夫误差下图的可匹配性

Zhirui Li[1],Keith D. Levin[2], Zhiang Zhao[3], 和 Vince Lyzinski[4]

[1]宾夕法尼亚大学统计与数据科学系[2]威斯康星大学麦迪逊分校统计系[3]肯塔基大学电气与计算机工程系[4]马里兰大学帕克分校数学系

2026年1月29日

摘要

我们研究由时变马尔可夫链噪声模型生成的一系列图的图匹配问题。我们的边灯游走误差模型是经典灯夫随机游走的一种变体，它通过边依赖噪声迭代地破坏图 G_0，从而生成一系列含噪图副本（G_t）。现有图匹配文献大多聚焦于边独立噪声情形下的匿名化阈值；本文则针对边依赖噪声情形，在匹配 G_0 和 G_t时建立了新的匿名化阈值。此外，我们还比较了这一匿名化阈值与马尔可夫链噪声模型的混合性质。我们证明，当 G_0 取自 Erd˝os–R´enyi 模型时，图匹配的匿名化阈值和边提灯人游走的混合时间均为 $\Theta(n^2 \log n)$ 阶。我们进一步证明，对于 G_0 更具结构性的模型（例如随机块模型），图匹配匿名化可在 $O(n^\alpha \log n)$ 时间内对某些$\alpha < 2$发生，这表明匿名化可能早于马尔可夫链噪声模型达到全局混合。通过大量模拟，我们在 Erd˝os–R´enyi 随机图和随机块模型随机图情形下验证了理论界，并在来自 Facebook 好友网络和一家欧洲研究机构电子邮件通信网络的真实世界数据集上探究了我们的发现。

1 引言

网络被广泛应用于各个科学领域，用来描述现实世界中复杂的关系。例如，研究《社会网络》时，节点代表用户，边表示友谊或关注关系 [53]。合著网络 [30] 和电子邮件交换网络[41] 已被证明有助于识别领军科研人员、知名机构以及潜在的学术欺诈行为。在神经科学领域，基于功能性磁共振成像扫描构建的网络 [69, 82] 已应用于分析连接模式 [28, 54], 以及

研究脑部相关疾病 [77, 81]。大型知识图谱在表示实体之间的概念关系方面变得越来越有用 [91]。

针对图集合的统计网络推断通常假设，观测到的图是由某些共享的潜在结构生成的带噪副本 [6, 44, 45, 60, 83]。基于这些观测，可以开展分类、假设检验等多种任务 [16, 22, 23, 33, 62, 74, 75, 78]。大多数此类方法的一个关键假设是，在进行推断之前，各网络中的顶点已经先验对齐 。若不满足这一条件，当顶点标签被打乱或带噪观测时，推断效果可能会下降 [65]。在这种情况下，通常会采用图匹配方法，尝试恢复各图中真实的节点标记，以便应用后续推断方法 [4, 6, 32, 39, 86]。

简而言之，图匹配的目标是找到两个（或多个）图的顶点之间尽可能最佳的对齐方式，以尽量减小逐边结构差异的测度。关于图匹配问题的详细讨论，包括其应用、近似方法、扩展和变体，请参阅综述 [19, 63, 87]。近年来，文献中涌现了大量致力于算法图匹配 [27, 34, 76, 80, 89, 90] 以及匹配理论的信息论极限[20, 21, 49, 85, 88]的论文。上述理论图匹配研究进展几乎都依赖于独立的边级噪声这一假设。然而，在蛋白质-蛋白质相互作用网络 [71], 、流行病学网络 [57], 、神经网络中的误差传播 [46], 以及时序网络 [8], 等实际应用中，考虑与时间或节点相关的噪声更为合理。为解决这一问题，本文关注具有空间或时间依赖性边噪声过程的图的可匹配性。

1.1 符号

对于任意 $n \in \mathbf{Z}^+$，记作 $[n] = \{1, 2, \ldots, n\}$。对于 $x \in \mathbf{R}$，记 $\lfloor x \rfloor$ 为 x的向下取整函数，即满足 $n \le x$ 和 $n+1 > x$ 的元素 $n \in \mathbf{Z}$；类似地，向上取整函数记作 $\lceil \cdot \rceil$。对于集合 S，记 $\binom{S}{j}$ 为其所有大小为 j 的子集组成的集合；也记 $\{0,1\}^{\binom{S}{j}}$为以$\binom{S}{j}$为索引的长度为 $\left|\binom{S}{j}\right|$ 的二进制序列集合。我们用 $\|\cdot\|_F$ 表示矩阵弗罗贝尼乌斯范数。我们采用渐近性的标准记号。特别地，对于两个函数 $f, g : \mathbf{Z}^+ \to \mathbf{R}^+$，若 $\lim_{n\to\infty} f(n)/g(n) = 0$，则记 $f(n) = o(g(n))$ 或$f \ll g$；若 $g(n) = o(f(n))$，则记 $f(n) = \omega(g(n))$ 或 $f \gg g$；若存在 $\exists C > 0$和 n_0使得 $\forall n \ge n_0$, $f(n) \le Cg(n)$，则记 $f(n) = O(g(n))$ 或 $f \lesssim g$；若 $g(n) = O(f(n))$，则记$f(n) = \Omega(g(n))$ 或 $f \gtrsim g$；若 $f(n) = O(g(n))$ 且$g(n) = O(f(n))$，则记 $f(n) = \Theta(g(n))$。

1.2 图匹配问题

本文所研究的图匹配问题定义如下。

定义 1. 考虑两个图 $G_1 = (V_1, E_1)$ 和 $G_2 = (V_2, E_2)$，其对应的邻接矩阵分别为 A 和 B。当 $|V_1| = |V_2| = n$时，图匹配问题

（图匹配问题）旨在寻找

$$\arg\min_{P\in\Pi_n}\|A-PBP^T\|_F=\arg\max_{P\in\Pi_n}\mathrm{Tr}\left(APBP^T\right), \tag{1}$$

其中 Π_n 表示所有 n乘n 置换矩阵的集合。

图匹配问题的目标是找到两个图中顶点的最佳可能对齐方式（以顶点标签的置换表示），从而最小化一种衡量边之间结构差异的测度（此处为二进制边差向量的 ℓ_2 范数）。在允许两个图为加权图、有向图且包含自环的一般情况下，上述表述等价于 NP 难的二次指派问题[51, 79]。尽管如此，对于某些特殊类别的图（例如平面图），仍存在高效的图匹配算法 [5, 66]。下文讨论图匹配问题时，由于加权图与方阵之间存在一一对应关系，我们将交替使用图及其邻接矩阵这两个说法。

图匹配问题有着悠久而丰富的历史。早期研究侧重于寻找图同构，其解使公式（1）中的目标函数等于零 [9, 61, 73]。然而，精确匹配的严格要求限制了图同构问题在现实场景中的适用性，因为这些场景中的图结构可能存在噪声或失真。为此，更一般的图匹配问题表述不再试图使目标函数降至零，而是将其最小化，从而允许两个图即便并非完全同构，仍可能具有有意义的结构相似性；更多背景信息请参见[19]。

在图同构问题中，存在真实对齐（即同构映射）这一概念，但在一般的图匹配问题中，未必存在这种对齐。为此，我们认为网络 G_1 和 G_2 的顶点集之间存在真实的、潜在的（即先验未知的）节点级对齐。这种对齐可以由数据驱动：在社交网络中，它可能表示同一用户在不同网络中的账户；在连接组学中，它可能表示对同一受试者进行不同扫描时对应的同一神经元区域；在蛋白质—蛋白质相互作用网络中，它可能表示两个不同相互作用网络中的同一蛋白质。这种对齐也可以由模型驱动，正如我们将在第 3.1 节中看到的那样，我们的随机图模型在 G_1 和 G_2 的边之间具有相关性，从而标识不同图中的顶点。本文中，为了指定真实对齐，我们使用“节点对齐”或“顶点对齐”来指代一对图 G_1 和 G_2（其邻接矩阵分别为 A 和 B），以及一个潜在对齐矩阵 P（不失一般性，我们通常令 P 为恒等对齐 I_n），使得对于所有 $i\in[n]$，A 中的节点索引 i 与 B 中的节点索引 $\sigma_P(i)$（其中 σ_P 是与置换矩阵 P 关联的置换函数）在每个图中都表示同一实体。无论采用何种算法，我们都将图的可匹配性定义为图匹配问题在信息论意义上能否恢复真实网络对齐；也就是说，若潜在对齐为 P^*，那么是否有 $\{P^*\}=\arg\max_{P\in\Pi_n}\mathrm{Tr}\left(APBP^T\right)$。

在合适的随机图模型中，已有文献指出，在某些情形下，可匹配性在理论上是可实现的 [85]，并且在实践中已得到实现 [26, 56]。如上所述，通常通过在网络之间的边引入相关性来确定图之间的节点对齐，而近期研究已确定这种边相关性存在一个尖锐阈值，超过该阈值，图匹配性便可行 [20, 21, 49, 85]。在这些模型中，边

在每个网络中， G_1 和 G_2 通常被假设为相互独立，而 G_1和 G_2之间的边则呈 ρ-相关，具体如下。对于每一对顶点 $\{u,v\} \in \binom{V}{2}$，有

$$\mathrm{corr}(\mathbb{1}\{\{u,v\} \in E(G_1)\}, \mathbb{1}\{\{u,v\} \in E(G_2)\}) = \rho.$$

此外，人们通常假设这组变量

$$\left\{\mathbb{1}\{\{u,v\} \in E(G_1)\}, \mathbb{1}\{\{w,z\} \in E(G_2)\}\right\}_{\{u,v\},\{w,z\} \in \binom{V}{2}}$$

除这种相关性外彼此独立。非正式地说，对于中等稠密图，要以高概率实现可匹配性，相关性的量级至少需要达到 $\sqrt{\log n/n}$ 。在这些尖锐阈值结果的设定中，不同的边对（例如 $\{u,v\}$ 和 $\{w,z\}$）无论是在单个图内还是在不同图之间，都是相互独立的。对于本文探讨的存在相关噪声结构的情形，目前只有少数研究建立了类似的阈值结果；例如参见 [6]。

1.3 随机图模型

为研究图的行为并进行统计推断，研究人员通常将研究范围限定在低秩图模型（即 E(A) 为低秩的模型）。在这些模型中，Erd˝os–R´enyi (ER) 图模型 [24] 最为简单，因为它假设所有边均以相同的概率独立地出现或缺失。

定义 2。设 n 为正整数，且 $p \in [0,1]$。若随机图 G 定义在 n 个顶点上，并服从参数为 (n,p) 的埃尔德什–雷尼模型（*ER*），记作$G \sim \mathrm{ER}(n,p)$，则对于任意一对满足 $i \neq j$的顶点 $i,j \in V(G)$，图 G 中 i 与 j 之间存在一条边的概率为 p，且该事件独立于所有其他边。

尽管简单，埃尔德什–雷尼模型已被证明是一个适用于数学分析的丰富模型，并且在该模型下已确立了图性质的各种阈值 [13, 24,31]。另请参见 [14]，其中介绍了“非齐次”埃尔德什–雷尼图：边彼此独立地存在或不存在，但其概率可能各不相同。

虽然我们关于马尔可夫（相关）噪声过程的混合与图匹配性之间相互作用的首个结果是在埃尔德什–雷尼模型中得到的，但更细致的结果需要具有更精细结构的图模型。为此，我们考虑埃尔德什–雷尼模型的一种扩展，该扩展假定边连接具有社区结构：随机块模型（SBM） [36]。

定义 3。设 n 和 K 为正整数，满足 $n \geq K$，且令 $\tau\colon [n] \to [K]$。给定一个对称矩阵 $\Lambda \in [0,1]^{K\times K}$，若对于任意一对节点 u、v，均有……，则一个含有 n 个顶点的图服从参数为 K、Λ、τ的随机块模型（*SBM*），记作 $G \sim$随机块模型（K、Λ、τ）。

$$\mathbb{P}(\textit{there is an edge between } u \textit{ and } v) = \Lambda_{\tau(u),\tau(v)}.$$

此外，事件的集合

$$\left\{\{there\ is\ an\ edge\ between\ u\ and\ v\}\right\}_{\{u,v\}\in\binom{V}{2}}$$

在给定 τ 的条件下相互独立。向量 τ 对 n 个顶点进行分区，将其划分为 K 个社群（即“块”），我们将每个 $k\in[K]$ 记为 $C_k=\{i:\ \tau(i)\ =k\}$，并定义向量 $\vec{n}=(n_1、\ n_2、\ .....、\ n_K)$ 表示社群大小 $n_k=|C_k|$。

本质上，随机块模型将 ER 随机图推广为一种允许边的存在概率根据顶点所属社群而有所不同的模型，而非让所有边都以相同的概率存在或不存在。学界已对随机块模型进行了深入研究，因为事实证明，它是一个丰富的模型，可用于从理论和算法层面探索网络中的社区检测性质；例如参见[1, 12, 38, 40, 52, 64, 72]。

1.4 MarkovError模型

给定图 G，可在 G 的顶点上定义随机游走：每一步都让行走者从一个顶点移动到另一个（可能相邻的）顶点，目标顶点以均匀随机方式选取 [43, 48]。在本文中，我们将把所有随机游走构造为具有适当定义的状态空间和转移矩阵的马尔可夫链。马尔可夫链的一个关键性质是，遍历链（即正常返、非周期且不可约的链）会收敛到其平稳分布 [43]。如果行走者正在扰乱给定图的边结构，且该链的平稳分布（定义在适当阶数的所有网络构成的空间上）足够平坦/均匀，那么随着链趋近平稳状态，单个顶点的身份最终会被匿名化。这一思想及其相互作用构成了下文理论发展的基础。

具体而言，我们构造一种灯夫式随机游走，其中灯位于图的边上，而非顶点上。在这种游走中，随机行走者每次经过一条边时，都会以一定概率打开或关闭这盏灯。将每一步之后提灯人的位置与所有边上的灯的状态结合起来考虑，我们便得到一个定义在空间 $V(G)\times\{1,0\}^{\binom{V(G)}{2}}$ 上的马尔可夫链。也就是说，该状态由随机行走者当前的位置，以及 G 的每条可能边上的灯的状态所共同决定，后者由 $\{1,0\}^{\binom{V(G)}{2}}$ 中的一个元素编码。关于类似构造的、灯位于顶点上的灯夫随机游走的详细讨论，参见 [43]。我们用于在图的边上引入误差的模型受到这些灯夫随机游走的启发，定义如下。

设 $G=(V,E)$ 为一个图，并考虑函数 $h_G:\ \binom{V}{2}\to\{0,1\}$，其定义如下

$$h_G(\{u,v\})=\begin{cases}1 & \text{if } \{u,v\}\in E\\ 0 & \text{otherwise.}\end{cases}$$

我们将称边 $e=\{i,j\}\in E$ 处于“开启”状态（即 $h(e)\ =1$），而将 $e\in\binom{V}{2}\setminus E$ 中的元素（即满足 $h(e)\ =\ 0$ 的元素）称为“关闭”状态。请注意，通过打开或关闭“灯”，

边提灯人会改变 E，从而改变图 G。我们用 G_0 表示初始图，用 G_t表示经过 t 步后的图，其对应的邻接矩阵分别为 A_0 和 A_t。

定义 4 (标准 edgelighter 随机游走**)**。设 $G_t = (V, E_t)$ 为时刻 t 图的状态，其中 $G_0 = G = (V, E_0 = E)$。边提灯人由一个时齐马尔可夫链 Z_t 建模，其状态空间为 $V \times \{0,1\}^{\binom{V}{2}}$。若时刻 t 边提灯人位于顶点 u，则边提灯人均匀随机地选择一个顶点 $v \in V$ 并移动到该顶点。若边提灯人在时刻 $t+1$选择了 $v = u$，则 $G_{t+1} = G_t$。若边提灯人在时刻 $t+1$选择顶点 $v \neq u$，则分两种情况讨论。

i. 若 $h_{G_t}(\{u,v\}) = 1$，则图通过 $G_{t+1} = (V, E_{t+1})$ 演化，其中

$$h_{G_{t+1}}(e) = \begin{cases} h_{G_t}(e) & \text{if } e \neq \{u,v\}; \\ X_t(1 - h_{G_t}(e)) + (1 - X_t)h_{G_t}(e) & \text{if } e = \{u,v\} \end{cases} \tag{2}$$

其中 $X_t \sim$伯努利分布(q_1) 与行走者和图历史 $(G_s)_{s \leq t}$相互独立。

ii. 如果 $h_{G_t}(\{u,v\}) = 0$，那么图通过 $G_{t+1} =$ （V， E_{t+1}）演化，其中

$$h_{G_{t+1}}(e) = \begin{cases} h_{G_t}(e) & \text{if } e \neq \{u,v\}; \\ Y_t(1 + h_{G_t}(e)) + (1 - Y_t)h_{G_t}(e) & \text{if } e = \{u,v\} \end{cases} \tag{3}$$

其中 $Y_t \sim$伯努利分布（q_2）与行走者和图历史 （G_s）$_{s \leq t}$相互独立。

简单地说，如果边提灯人 *traverses* 一条“开启”边（相应地，一条“关闭”边），那么边提灯人以 q_1 的概率（相应地，以 q_2 的概率）独立地将该边从开启切换为关闭（相应地，从关闭切换为开启），并以 $1 - q_1$ 的概率（相应地，以 $1 - q_2$ 的概率）保持该边开启（相应地，保持关闭）。

我们在此指出，上述定义的马尔可夫链是非周期、不可约的，并且关于测度 π°（（u， c））$\propto {}_{n}^{1}q^{\#\text{中开启边的}}{}_{2}^{}{}^{c}q^{\#\text{中关闭边的}}{}_{1}^{}{}^{c}$可逆。前两个断言是显然的。可逆性由以下论证得出。设（u， c）和（v， c'）是 $V \times \{1,0\}$ （$\binom{V}{2}$）中的两个状态，使得 c和 c'在所有可能的边上都相同，除了可能在 $\{u,v\}$ 处不同（其中 $u \neq v$）。注意，如果 $c = c'$，那么

$$\mathbb{P}_{(u,c)}(Z_1 = (v,c')) = \mathbb{P}_{(v,c')}(Z_1 = (u,c)) = \begin{cases} \frac{1}{n}(1 - q_1) \text{ if } c(\{u,v\}) = 1; \\ \frac{1}{n}(1 - q_2) \text{ if } c(\{u,v\}) = 0. \end{cases}$$

如果 $c(\{u,v\}) \neq c'(\{u,v\})$，那么我们有

情形 1： $c(\{u,v\}) = 1$，在这种情况下

$$\mathbb{P}_{(u,c)}[Z_1 = (v,c')] = q_1/n \text{ and } \mathbb{P}_{(v,c')}[Z_1 = (u,c)] = q_2/n.$$

情形 2： $c(\{u,v\}) = 0$，在这种情况下

$$\mathbb{P}_{(u,c)}[Z_1 = (v,c')] = q_2/n \text{ and } \mathbb{P}_{(v,c')}[Z_1 = (u,c)] = q_1/n.$$

设 m 为 c 中 1 的数量（即 c中的边数），r 为 0 的数量（即非边的数量）。在情形 1 中，

$$\pi^{\circ}(\,(u,c)\,)\mathbb{P}_{(u,c)}(Z_1=(v,c'))\propto\frac{1}{n}q_2^m q_1^r\frac{1}{n}q_1,$$
$$\pi^{\circ}(\,(v,c')\,)\mathbb{P}_{(v,c')}(Z_1=(u,c))\propto\frac{1}{n}q_2^{m-1}q_1^{r+1}\frac{1}{n}q_2,$$

两者相等。情形 2 的证明类似，细节从略。由此立即可得 (Z_t) 的遍历性。

备注 1. 若令 L_t 表示边灯随机游走在时刻 t 所处的位置，则过程 $(Y_t=\{L_{2t},L_{2t+1}\})_{i=1}^{\infty}$（注意这里是无序对）等价于在顶点集 $\binom{V}{2}\cup\{\{v,v\}$ 上的完全图（含自环）上的随机游走，其中 $v\in V\}$.。稍后我们将特别关注该游走的覆盖时间，而优惠券收集问题中的经典结果 [25] 将使我们能够把当前工作与关于边上独立同分布噪声下可匹配性的标准结果联系起来，相关讨论见 [85]；详情请参见第 3.1 节。这里指出，过程 (X_t,C_t)——其中 $X_t=(L_t,L_{t+1})$，而 C_t 是 $\{1,0\}^{\binom{V}{2}}$ 在时刻 t 的灯构型——类似于经典的灯夫游走（采用[43, 第 19] 章中的术语），该游走定义在线图 D_V 上，其中 D_V 是顶点集为 V 的完全有向图。当前模型的关键区别在于，我们不会像经典灯夫游走那样，在访问边和离开边时都对其随机化；此外，我们的马尔可夫链也不是惰性的。

2 混合与匹配

本研究旨在理解我们的边提灯人游走过程所引入的序贯误差结构如何影响图匹配性。为此，考虑将边提灯人游走在时间步 t 时的图 G_t 与原始图 G_0 对齐的问题。具体而言，我们旨在了解行走者在匹配无法恢复大量正确顶点对应关系之前，最多可以行走多少步。为了形式化定义匿名化时间，我们考虑如下内容：对于 $\beta\in(0,1)$，我们通过以下方式定义 β-匿名化时间 t^{β}_{anon}:

$$t^{\beta}_{\text{anon}}=\min\left\{t\in\mathbb{Z}^{\geq 0}\text{ s.t. }\arg\min_{P\in\Pi_n}\|G_tP-PG_0\|_F\subset\bigcup_{k>n^{\beta}}\Pi_{n,k}\right\},$$

其中 $\Pi_{n,\,k}$ 是恰好打乱 k 个顶点标签的置换集合，适用于 $0<k\leq n$（注意，这也将在下文的定义 β中通过定义 6 的 β-匿名化予以形式化）。集合 $\Pi_{n,\,k}$ 中的置换在组合数学文献中称为部分错排[70]。稍后我们会对此作更精确的说明，但直观而言，在此时刻，t^{β}_{anon}，图 G_0 中存在的大部分结构已被噪声过程有效抹去。同时，我们还希望将含噪图的这一匿名化时间与上一节定义的底层马尔可夫链（Z_t）的混合时间联系起来。

对于遍历马尔可夫链，现代马尔可夫链理论的一个分支致力于分析其收敛至平稳分布的速度，由此引出了马尔可夫链混合时间的定义 [3, 43]。

定义 5. 设(X_n) 是状态空间为 S 、一步转移概率函数为 P: $S \times S \to [0,1]$ 、平稳分布为 π 的离散时间遍历马尔可夫链。该马尔可夫链的全变差混合时间定义为

$$t_m = \min\left\{t \geq 0 : \max_{x \in S} \left\|P^t(x,\cdot) - \pi\right\|_{TV} < \frac{1}{4}\right\}.$$

直观地说，混合时间刻画了马尔可夫链的分布趋近其平稳分布所需的步数。我们注意到，$1/4$的取值是按照惯例选定的——它可以替换为任意适当小的常数；事实上，如果我们定义$t_m(\epsilon) = \min\{t \geq 0 : \max_{x \in S} \|P^t(x,\cdot) - \pi\|_{TV} < \epsilon\}$，那么 $t_m(\epsilon) \leq \lceil \log \epsilon^{-1} \rceil t_m$。混合时间分析广泛应用于多个领域，包括统计物理、随机化算法和概率组合学 [10, 11, 42, 43, 55, 84]。因此，人们开发了多种方法来分析混合时间，例如谱隙分析和耦合方法 [3]。

在分析随机网络上的随机游走时，混合时间分析也已成为一项重要任务。例如， [68] 给出了 Erd˝os-R´enyi 图上时间齐次马尔可夫链的混合时间，而 [37]则分析了如何利用特定的图结构来实现更快的混合。基于此，我们将考虑 Erd˝os-R´enyi 图和随机块模型随机图上的边轻化图的混合时间分析（即， G_0 是从 Erd˝os-R´enyi 图或随机块模型中抽取的）。我们在此推测，当经过的时间超过马尔可夫链的混合时间（Z_t）后，边提灯人已将原始边信号 G_0中的大部分随机化。事实上，我们很快就会准确阐明这一直觉：边提灯人混合后，边概率已充分趋近平稳状态，从而导致可匹配性损失。

为了刻画这种可匹配性损失，我们考虑以下定义，旨在描述通过将打乱后的 G_t 与 G_0进行匹配来恢复其顶点标签的能力。为此，我们寻求关于 t的条件，使得我们仍能正确地将 G_t与 G_0匹配。要求“完美匹配”恢复所有 n顶点的真实对应关系过于严格，因此这里采用一种较宽松的可匹配性概念。

定义 6 比率保持匹配。 设 $G_1 = (V_1, E_1)$ 和 $G_2 = (V_2, E_2)$ 是两个顶点对齐图，其邻接矩阵分别为 A 和 B。设 $\alpha, \beta \in (0,1)$ 为常数。我们接着定义

1. α**-可匹配性**。 如果 G_1 和 G_2 是 α-可匹配的，则

$$\mathbb{P}\left[\exists P \in \bigcup_{k \geq n^\alpha} \Pi_{n,k} \;\; s.t. \;\; \mathrm{Tr}\left(APBP^T\right) \geq \mathrm{Tr}(AB)\right] = o(1).$$

换言之，如果找到一个置换的概率渐近趋于零，而该置换至少打乱 n^α 个节点，且其匹配效果不逊于真实恒等映射（以图匹配问题的目标函数衡量），那么我们就称图 G_1 和 G_2 是 α-可匹配的。

2. β-匿名化。我们称 G_1 和 G_2 是 β-匿名化的，当且仅当

$$\mathbb{P}\left[\arg\max_{P \in \Pi_n} \mathrm{Tr}\left(APBP^T\right) \subset \bigcup_{k=0}^{n^\beta} \Pi_{n,k}\right] = o(1).$$

换言之，如果两个图 G_1 和 G_2 之间的最优匹配（以 *GMP* 目标函数衡量）至多打乱 n^β 个节点的概率渐近趋于零，则称它们是 β-匿名化的。换言之，以高概率，最优匹配会打乱超过 n^β 个节点。

3 Erd˝os-R´enyi 设置下的混合与匹配

我们旨在探讨上述 edgelighter 马尔可夫链的混合时间与初始图的可匹配性之间的关系 G_0 以及图 G_t 在时刻t；回忆一下， A_0 和 A_t 分别表示它们对应的邻接矩阵。我们首先考虑定义4中的标准 edgelighter 随机游走。

注意，令 L_t 表示灯夫在时刻 t访问的顶点，则序列$\{Y_t = \{L_{2t}, L_{2t+1}\}\}_{t=0}^{\infty}$构成状态空间 $S = \binom{V}{2} \cup \{\{v, v\}$ 上的马尔可夫链： $v \in V\}$；参见备注1。令 t_{cover} 为该马尔可夫链（Y_t）的覆盖时间，即，

$$\tau_{\text{cover}} = \min\{t : S \subset \cup_{\ell \le t} Y_\ell\}; \quad t_{\text{cover}} = \max_{s \in S} \mathbb{E}_s(\tau_{\text{cover}}). \tag{4}$$

其中 $\mathbf{E}_s(\cdot)$ 表示灯夫游走从状态 s开始时的期望。注意，简单的优惠券收集者渐近性告诉我们，该马尔可夫链的覆盖时间 t_{cover} 为 $\Theta(n^2 \log n)$ 阶，并且 τ_{cover} 以高概率为 $\Theta(n^2 \log n)$ 阶。若选择 q_1 和 q_2 使得 $q_1 = 1 - q_2$，那么对于 $t \ge \tau_{\text{cover}}$，我们有 G_t 服从 $\mathbf{ER}(n, q_2)$ 分布，且 G_t 与 G_0 独立。在这种情况下，显然

$$\mathbb{P}\left(I_n \in \arg\min_{P \in \Pi_n} \|A_0 - P A_t P^T\|_F \,\middle|\, t \ge \tau_{\text{cover}} \right)$$

对于所有 $t > \tau_{\text{cover}}$，都与 $\mathbf{1}$ 保持一定距离，因为到此时，任何置换成为最优解的可能性都相同。

接下来，我们的第一个引理将 (Z_t) 的混合时间（记为 t_m°）与 (X_t) 的一个适当定义的覆盖时间联系起来，其中 $X_t = (L_t, L_{t+1})$。定义 $\tau_{\text{cover}}^{(u)}$ 如下：

$$\tau_{\text{cover}}^{(u)} = \min\left\{ t : \binom{V}{2} \subset \bigcup_{\ell \le t-1} \{L_\ell, L_{\ell+1}\} \right\},$$

并定义 $t_{\text{cover}}^{(u)} = \max_x \mathbb{E}_x \tau_{\text{cover}}^{(u)}$。注意，显然 $\tau^{(u)\text{覆盖}} \le \tau_{\text{cover}}$，其中 τ_{cover}的定义见式 4；以下引理的证明见附录7.1。

引理 1。 采用上述记号，设 $q_1 = 1 - q_2$ 属于定义 4中的 *edgelighter* 游走。于是有 $\frac{1}{8} t_{cover}^{(u)} \le t_m^\circ \le 11 t_{cover}^{(u)}$。

简单的优惠券收集者渐近（证明见附录7.2 ）表明，对于足够大的 n,

$$\frac{1}{2} n^2 \log n \le t_{\text{cover}}^{(u)} \le \frac{5}{2} n^2 \log n, \tag{5}$$

因此 $t_m^\circ = \Theta(n^2 \log n)$，这里回顾一下， t_m° 是 Z_t的混合时间。

现在，我们将重点转向探究 G_0 与边提灯人游走的模型，在这些模型中，信号恢复在时间 t_{cover} 之前是可能的；同时找出信号恢复在时间 t_{cover}之前失败的情形。我们首先考虑 G_0 服从 ER 随机图分布的情况。

3.1 Erd˝os–R´enyi 图上的标准 edgelighter 随机游走

我们首先考虑 ER 图 $G_0 = (V, E) \sim \text{ER}(n, p)$ 上的标准 edgelighter 随机游走。我们假设边提灯人的初始起始顶点是从 V中均匀随机选取的。令 G_t 表示经过 t 步 edgelighter 游走后的图；在这里，我们考虑参数 $q_1 = (1-p)$， $q_2 = p$，使得 $G_t \sim$ 也服从 $\text{ER}(n, p)$ 分布。

为了计算此设定下 G_t （其邻接矩阵为 A_t）与 G_0（其邻接矩阵为 A_0）之间的边级相关性，定义

$$\mathfrak{p}_{t,u,v} = \mathbb{P}\Big(\{u,v\} \notin \cup_{s\le t-1}\{L_s, L_{s+1}\}\Big);$$

也就是说，$\mathfrak{p}_{t,\,u,\,v}$ 是边 $\{u,v\}$ 截至时刻 t尚未被边提灯人经过的概率。在附录7.2中，我们证明了

$$\mathfrak{p}_{t,u,v} = \exp\left\{-\Theta\left(\frac{t}{n^2}\right)\right\} \tag{6}$$

边 $\{u,v\}$ 被经过后，其开/关状态与初始开/关状态无关，因此我们有

$$\text{Corr}(A_{0,u,v}, A_{t,u,v}) = \frac{\mathbb{E}(A_{0,u,v}A_{t,u,v}) - p^2}{p(1-p)} = \frac{p\mathfrak{p}_{t,u,v} + p^2(1-\mathfrak{p}_{t,u,v}) - p^2}{p(1-p)} = \mathfrak{p}_{t,u,v}$$

不过我们注意到，与传统的相关 Erd˝os-R´enyi 模型不同，在该模型中，不同顶点对之间的边在不同图中相互独立，而边提灯人会在网络 G_0 和 G_t之间引入一种微妙的依赖结构。例如，在时刻 t，如果 G_t 中有 t 条边的开/关状态与 G_0中的对应边相反，那么 G_t 中其余的边必定与 G_0中的对应边完全相同。尽管如此，我们猜测这种依赖会随时间减弱，并推测可匹配性相变发生在逐边相关性为 $\sqrt{n^{-1}\log n}$ 的量级（与相关 Erd˝os-R´enyi 模型[49, 50, 85]中的量级相同），这意味着 $t = \Omega(n^2 \log n)$ 足以使这些图实现有效匿名化；并且，如果 $t = O(n^2 \log n)$，则可匹配性得以保持。

采用定义 6中给出的匹配，接下来我们将给出一系列结果，对前文猜想的可匹配性阈值作出（至少部分）肯定回答。在以下定理中，我们考虑 $G_0 \sim ER(n,p)$，并考虑 $1 - q_1 = q_2 = p$。在每一步中，我们再次将 G_t 与 G_0进行匹配，并得到以下定理（证明见附录 7.3）。

定理 1。令 $\{G_t\}_{t\in\mathbb{N}}$ 如上所述，并满足 $1 - q_1 = q_2 = p$。令 α、$c > 0$ 为满足 $\alpha > 5c$的固定常数。若 $t \le cn^2$ 对数 n，则 G_0 和 G_t 是 α-可匹配的；具体而言，

$$\mathbb{P}\left[\exists P \in \bigcup_{k\ge n^\alpha} \Pi_{n,k} \ s.t. \ \text{Tr}\big(A_0 P A_t P^T\big) \ge \text{Tr}(A_0 A_t)\right] \le 2\exp\left\{-\Omega\left(\frac{n^{2\alpha-5c}}{\log n}\right)\right\}.$$

使用上述记号，对于匹配匿名化界，我们考虑时间 $\tau_{\text{cover}}^{(u)}$，因为在此时间之后， G_t 实际上是独立于 n、 p的 ER(G_0 ，且任意置换成为最优置换的可能性均相等。如注 1所述，（无序）对 $(\{L_{2i}, L_{2i+1}\})_{i=1}^{\infty}$彼此独立，并在 $S = \binom{V}{2} \cup \{\{v, v\}$ 上均匀分布： $v \in V\}$。令 τ_{cover} 如式 4 所示，根据标准的优惠券收集者渐近性可得（当 $|S| \leq n^2$ ）

$$\mathbb{P}\left[\tau_{\text{cover}}^{(u)} \geq 4n^2 \log(n)\right] \leq \mathbb{P}\left[\tau_{\text{cover}} \geq 4n^2 \log(n)\right] \leq n^2 e^{-4 \log n} = n^{-2}. \tag{7}$$

接下来给出以下定理。证明见附录7.4。

定理2. 设 $\{G_t\}_{t \in \mathbb{N}}$ 如定理*1* 中所述，且 $1 - q_1 = q_2 = p$。若 $t \geq 4n^2 \log n$，则对于任意 $\beta \in (0,1)$， G_0 和 G_t 均为 β-匿名化；具体而言：

$$\mathbb{P}\left[\arg\max_{Q \in \Pi_n} \operatorname{Tr}\left(A_0 Q A_t Q^T\right) \subset \bigcup_{k=0}^{n^\beta} \Pi_{n,k}\right] \leq n^{-2} + (2n^\beta)^{1-n/(2n^\beta+1)}$$

定理1和2 表明，对于标准边轻化过程， A_t与 A_0 之间的去匿名化发生所需的时间，与 (Z_t) 的混合时间同阶，即需要 $t = \Theta(n^2 \log n)$ 步。

备注 2. 注意，若1和2 ，则定理 $1 - q_1 = q_2 = q$ 对于任意 $q \in (0,1)$，在适当调整常数后仍成立。为与 $q = p$ ER(n, p) 模型在 $\tau_{\text{cover}}^{(u)}$前后的参数保持一致，我们采用。

4 结构化图的混合前（部分）匿名化

我们观察到，对于 G_0 ，若其采样自简单 ER 模型，则噪声过程（Z_t）的混合时间与匿名化时间一致。本文旨在找出匹配失效的模型（即在噪声混合网络分布之前，图已按照定义 6 得到充分匿名化的模型）。为此，我们考虑一种结构更复杂的图分布：随机块模型（ SBM）。我们将考察的具体模型如下：

$$\text{SBM}\left(K_n, \Lambda_n = [\Lambda_{i,j;n}], \tau_n\right) \subseteq \mathcal{G}_n,$$

在此随机块模型中，我们将 edgelighter 游走定义如下

定义 7 (分块边灯笼人游走)。 令 $G_t = (V, E_t)$ 表示图在时刻 t 的状态，并假设 G_0 的顶点集被划分为 K 个社群 $B_1, B_2 \ldots, B_K \subseteq V$（例如， G_0 可以从上文指定的随机块模型中抽样得到）。这里的边提灯人仍建模为一个时间齐次马尔可夫链(W_t)，但此时其状态空间为 $\mathcal{C} \times V \times \{1,0\}^{\binom{V}{2}}$，其中 $\mathcal{C} = [K]$ 是图中所有社群标签的集合。如果提灯人位于社群 i （即处于 B_i）且位于顶点 $u \in B_i$，那么下一步它以相等的概率（即各为 *1/2*）留在或离开社群 B_i 。

i. 如果边提灯人选择留在社群 B_i，则提灯人均匀随机选择一个顶点$v \in B_i$，并移动到 v。如果 $v = u$，则 $G_{t+1} = G_t$，不进行任何操作。如果边提灯人在时刻 $t+1$ 移动到顶点 $v \neq u$，则分两种情况讨论。如果 $h_{G_t}(\{u, v\}) = 1$，那么图通过 $G_{t+1} = (V, E_{t+1})$ 演化，其中

$$h_{G_{t+1}}(e) = \begin{cases} h_{G_t}(e) & \text{if } e \neq \{u, v\}; \\ X_t(1 - h_{G_t}(e)) + (1 - X_t)h_{G_t}(e) & \text{if } e = \{u, v\} \end{cases} \tag{8}$$

其中，$X_t \sim$ 伯努利分布$(q_{i;1})$ 与边提灯人游走以及$(G_s)_{s \leq t}$相互独立。如果 $h_{G_t}(\{u, v\}) = 0$，则图通过 $G_{t+1} = (V, E_{t+1})$ 演化，其中

$$h_{G_{t+1}}(e) = \begin{cases} h_{G_t}(e) & \text{if } e \neq \{u, v\}; \\ Y_t(1 + h_{G_t}(e)) + (1 - Y_t)h_{G_t}(e) & \text{if } e = \{u, v\} \end{cases} \tag{9}$$

其中，$Y_t \sim$伯努利分布$(q_{i;2})$ 与边提灯人游走以及$(G_s)_{s \leq t}$相互独立。

ii. 如果边提灯人选择离开 B_i，那么提灯人首先从所有 $K_n - 1$ 非B_i社群中均匀随机选择一个新的社群$B_j \neq B_i$。然后，边提灯人独立地在 B_i 和 B_j 之间均匀随机选择一条边 e_1，并在 B_i 和 B_j 之间均匀随机选择一条非边 e_2，并设置$h_{G_{t+1}}(e_1) = 0$ 和$h_{G_{t+1}}(e_2) = 1$。随后，边提灯人从 B_j中的节点里均匀随机选择一个节点 $w \in B_j$，并在时间 $t+1$移动到 w。

我们可以将这个边提灯人想象为在具有社群的图上进行分层游走。在社区层面，边提灯人在各个社区之间移动，并将所选社区之间的边均匀随机化。在社区内部，边提灯人的行为则类似于标准 *edgelighter* 随机游走。

与简单的边提灯人链一样，块边缘轻量器链是非周期、不可约的（在受限状态空间上，任意两个社区之间的边数均固定为 G_0 中的值），且可逆。这里，平稳分布 $\pi^\bullet$ 定义如下：

$$\pi^\bullet(B_i, u, \vec{c}) \propto \frac{1}{K}\frac{1}{|B_i|}\prod_j q_{j;2}^{m_{j,\vec{c}}} q_{j;1}^{r_{j,\vec{c}}} \frac{1}{\prod_{i,j}\binom{n_i n_j}{m_{ij}}},$$

其中，K 是社区数，$u \in B_i$，且 $\vec{c}$ 是边配置，其中 m_i，$\vec{c}$ 是 B_i 与 r_i 中的顶点之间现有的 1 的数量，$\vec{c}$ 是 B_i 中现有的 0 的数量，n_1 是 B_i 中的顶点数，m_{ij} 是 B_i 与 B_j 之间的边数；关于这种平稳分布形式的证明，请参见附录 7.6。定义了分块边灯笼人游走后，我们得到以下结果：在块边缘轻量器噪声模型完全混合之前，对图中一个社区进行局部 β-匿名化。若社区 B_i 的大小为 n_i，则称其在时间 t 已达到 β-匿名化，当且仅当满足以下条件（其中 Π_{B_i}，β 是一组置换，其中 B_i 中至多有 n_i^β 个顶点被打乱）

$$\mathbb{P}\left(\arg\max_{P \in \Pi_n} \mathrm{Tr}(A_0 P A_t P^T) \subset \Pi_{B_i,\beta}\right) = o(1).$$

请注意，定理 3 的证明见附录 7.5。

定理 3. 假设存在常数 b_1、b_2、b_3、$b_4 > 0$ 和 $0 < a_1 < a_2 < 1$，并且不失一般性地假设 $n_1 =$ 最小值 n_i,

$$n_1 = b_1 n^{a_1}; \quad n^* := \max_i n_i = b_2 n; \quad \Lambda_{1,j;n} \leq b_3 \frac{n^{a_2}}{n_j n_1} \text{ for all } j \neq 1; \quad K_n = b_4;$$

$$\Lambda_{i,j;n} \geq 5\frac{\log n}{n_i n_j} \text{ for all } i, j \text{ with } i \neq j; \quad \Lambda_{i,i;n} \geq 5\frac{\log n}{\binom{n_i}{2}} \text{ for all } i$$

考虑上文定义的随机块模型灯夫，其中对所有 i 都有 $1 - q_{i;1} = q_{i;2} = q$。定义事件

$$\mathcal{E}_{t,\beta} = \left\{ \underset{P \in \Pi_n}{arg\ max} \operatorname{Tr}(A_0 P A_t P^T) \subset \Pi_{B_1,\beta} \right\},$$

回想一下，$\Pi_{B_1,\ \beta}$ 是将社区 B_1 中至少一定比例的β 标签打乱的置换矩阵集合；令 $N = n^{1+a_2}\log^{3/2}(n)$。那么，对于某个适当的常数 $C > 0$，有：

$$\begin{aligned}\mathbb{P}(\mathcal{E}_{N,\beta}) \leq O\left((2n_1^{\beta})^{-n_1/(2n_1^{\beta}+1)}\right) + O\left((n_1 n^*)^{-Cn^{a_2}}\right) + O(n^{2a_2}\log(n)\, b_4^{-n^{(1-a_2)/2}}) \\ + O(n^{a_2-1}/2)) + e^{-\Omega(n^{a_1}\log^{1/2}(n))}\end{aligned}$$

我们将网络中至少一个社区被 β-匿名化（在我们的设定中为社区1）的阶段称为网络局部匿名化的时刻。定理 3 结合下文的引理 2 表明，在随机块模型上的分块边提灯人满足温和假设时（关键在于存在一个规模小于最大社群的社区），匹配会先在局部被破坏，之后马尔可夫噪声才会全局混合。注意，定理 1不适用于此处，因为社团结构会导致在社区内部或社区之间行走的概率不同。以下引理的证明见附录 7.7，它完成了这一论证：

引理 2。 在定理3的假设下，令 $t_m^{\bullet}$ 为分块边灯笼人游走的混合时间。于是有 $t_m^{\bullet} = \Omega(n^2)$。

混合时间为 $t_m^{\bullet} = \Omega(n^2)$，因此经过 $n^{1+a_2}\log^{3/2}(n)$ 步（足以实现局部 β-匿名化）后，马尔可夫链（W_t）尚未实现全局混合。

定理 β证明的局部匿名化是定理3中所证结果的推论，其原因在于在某种意义上，最小社区已实现局部混合：确实，在实现局部 β-匿名化时，该最小社区内部及跨越其边界的边已充分混合（即，这些边的平稳概率接近相应的平稳概率）。最后，尽管对于分块边灯笼人游走，局部匹配早于噪声的混合时间被破坏，定理3 仅给出了无法实现精确匹配的条件，正如 [85]所定义的。根据定理3，经过 $\Theta(n^{1+a_2}\log(n))$步后，我们只能保证常数数量的社区（至少 1 个）实现匿名化，而非整个网络。我们推测，要确保全局 β-匿名化，仍需 $\Omega(n^2\log(n))$ 时间；对此类界的进一步分析将留待未来研究。

注 3。 在定理3中，随机块模型中的参数假设可以进一步收紧。实质上，这些假设需要确保最小社群 n_1 的阶低于

最大社群，并且连接最小社群与图中其余部分的边不能太多。由于我们从启发式角度关注这个例子，因此本文不探讨必要且充分条件。

5 实验

为了进一步探究我们的理论，我们开展了两组实验。第一组中，边轻化过程在合成随机基图上游走（基图为 Erd˝os-R´enyi 图或随机块模型基图）；第二组中，边提灯人在真实数据基网络上游走 G_0。在这些模拟中，我们估算了适当的匿名化时间和 edgelighter 游走的混合时间，旨在表明匿名化大致发生在混合时刻。需要注意的是，由于精确的混合时间难以量化，我们在标准 edgelighter 游走中使用引理1 和2 ，使混合时间 t_m° 与该游走的覆盖时间 t_{cover} （至少渐近地）相等；因此，我们调整后的目标是表明匿名化大致发生在覆盖时刻。在随机块模型/块设置中，覆盖时间为混合时间提供下界。

为解决精确图匹配问题的计算难处理性，我们采用 [29]提出的种子图匹配（SGM）算法，具体如下。为加快随机块模型游走和真实数据上的可匹配性实验，我们将匹配算法初始化为真实值——单位矩阵。请注意，发生匿名化时，算法的优化器应偏离真实值。虽然真实值为局部最优可能会阻止这一变化，但这种初始化为评估真实值的最优性提供了良好的启发式方法，同时有效节省了计算资源。SGM 算法 [29, 67] 还利用种子顶点（即对应关系事先已知且在整个匹配过程中固定的顶点）来提升匹配性能。在实验中，我们从节点集合中均匀随机地选取种子，算法则确保匹配结果将这些种子节点映射到自身。除非另有说明，我们均匀随机地选取 5% 的节点作为种子。

5.1 ER图和随机块模型图上的模拟边轻化器游走

我们首先考察定义 4 中标准边轻化器游走模型在 ER 图上的表现。我们采样了具有 $p = 0.5$ 以及节点数 $n = 49$、100、144、225、324 和 729 的 ER 初始图；其中 $q_1 = q_2 = 1/2$。对于每个图，我们执行标准边轻化器游走。对于较小的图（$n = 49$、100、144），我们在每次移动后检查匹配正确性。为节省较大图（$n = 225$、324、729）的计算资源，我们在固定步数后检查匹配，具体而言，对于 $s_n = 3$、324 和 729，分别在 30 步和 300 步后进行检查。在图1 的上面板中，我们绘制了 $n = 100$、225 和 $n = 729$ 的匹配正确性与步数之间的关系（蓝色）。其他节点数取值的结果见附录 7.8.1。此外，我们还用红色绘制了这些实验中边的覆盖率与步数之间的关系。在图1的下面板中，我们展示了发生 0.5-匿名化之前所需步数与节点数之间的双对数图，并附上拟合回归线。其他双对数尺度下的图还包括

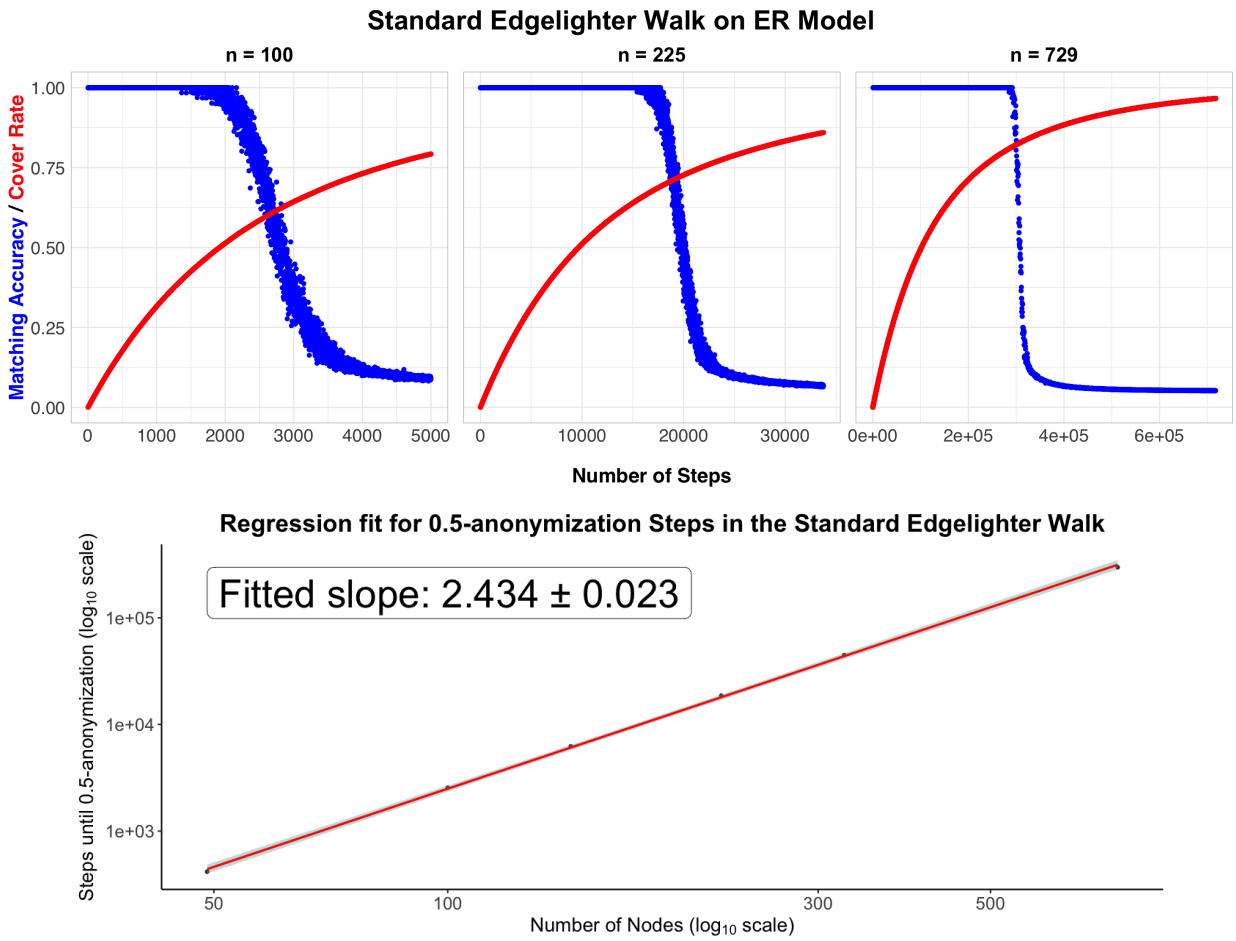

图1：在上图中，我们以蓝色绘制了 $n = 100$（左）、 $n = 225$ （中）和 $n = 729$ （右）的匹配正确性与步数之间的关系。对于所有这些图，我们进一步以红色绘制了边的覆盖率与步数之间的关系。在下图中，我们以双对数尺度绘制了发生 0.5-匿名化之前所需步数关于节点数 n 的回归图。

发生 0.25-匿名化以及发生 0.75-匿名化之前所需的步数；对于同一次运行，相关结果见附录 7.8.1。拟合直线的斜率略高于 2，这与我们的理论结果一致，即需要 $\Theta(n^2 \log n$ 步才能破坏可匹配性。

接下来，我们考虑从随机块模型（SBM）中采样得到的结构更为规整的分块 edgelighter on G_0，以说明一种匿名化发生在混合之前的情形。这里，我们从以下随机块模型中采样图:

$$\mathrm{SBM}\left(K=\left\lfloor\frac{n-n^{1/4}-n^{4/3}}{n^{2/3}}\right\rfloor,\ \Lambda=\frac{\log(n)}{n^{3/4}}J+\mathrm{diag}(1/2),\ \tau=\tau_n\right)$$

其中 τ_n 由块大小向量 $\vec{n} = (n_1$、 $\cdots$、 $n_K) = (n^{1/4}$、 $n^{2/3} \cdot \mathbf{1}_{K-2}^T$、 $n^{3/4})$ 根据 $\tau_n(j)$ 表示 $= \sum_{k=0}^{K-1} \mathbb{1}_{\{j \geq \sum_{i=1}^{k} n_i\}}$ 其中 $n_0 = 0$。这些参数的选择旨在反映定理 3中的关键条件，即社区不平衡以及.....的稀疏性

社区 1 与图中其余部分之间的连接；请注意，在这些条件下可以证明定理 3 的类似结论。实验中，我们选择节点数量$n = 81$、256 和 625，并在这些采样网络上执行 SBM 边轻化游走。对于 $n = 81$，每一步后进行匹配；对于 $n = 256$，每 90 步后进行匹配；对于 $n = 625$，每 2100 步后进行匹配。此外，请记得我们将匹配算法初始化为单位矩阵。

在图2中，我们展示了 $n = 256$ 个节点情况下的部分选定图； $n = 256$ 情况下其他社区的图，以及 $n = 81$ 和 $n = 625$情况下的其他图，可见附录7.8.2。在图2的上方面板中，我们以蓝色展示了整个网络以及三个选定社区的匹配正确性随步数的变化：社区1（规模最小的社区）、社区2（随机选取的一个、规模为 $n^{2/3}$的社区），以及社区7（规模最大的社区）。在所有这些图中，我们还以红色展示了边覆盖率随步数的变化。

我们观察到，规模最小的社区匿名化明显更快，这与我们提出的理论框架高度吻合。此外，与其他社区相比，规模最大的社区匿名化速度较慢。还需注意的是，将此随机块模型边提灯人设置与 ER 图上的标准边轻化设置进行比较时，可匹配性消失的转变似乎不那么陡峭。我们还在图2 的下方面板中绘制了达到 0。5 匿名化所需步数与节点数之间的双对数图，并在这些点上叠加拟合的回归线。同一分块边提灯人运行中，达到 0。25 匿名化和 0。75 匿名化所需步数的其他双对数尺度图见附录7.8.2。如图所示，回归线的斜率小于 2，且比 ER 网络上的标准 edgelighter 随机游走所得斜率约小 0.5。这个数值与我们在第4节理论框架所提出的结果一致。

5.2 真实数据实验

为了在真实世界网络上验证所提出的理论结果，我们将标准边轻化模型应用于一个来自 [53]的友谊网络。我们从https://snap.stanford.edu/data/ego-Facebook.html下载了该网络，并提取了包含第 1921 至 2640 个节点的诱导子图（共 $n = 720$ 个节点）。图3左侧面板展示了相应邻接矩阵的图示。我们在选定的网络上运行标准 edgelighter 随机游走，共进行 900,000 步，每隔 150 步与原始网络进行一次匹配。edgelighter 游走引入的噪声可以解释为个人结交新朋友，或与近期没有互动的人失去联系。我们还追踪每一步之后的覆盖率，其定义为

$$\frac{\#\text{ distinct edges traversed after the step}}{\binom{720}{2}}.$$

在图 3 的右侧面板中，我们以蓝色绘制匹配准确率与步数的曲线，同时以红色绘制覆盖率与步数的曲线。结果显示出与我们在 Erd˝os–R´enyi 图上观察到的标准 edgelighter 随机游走相似的模式，如图 1 所示。

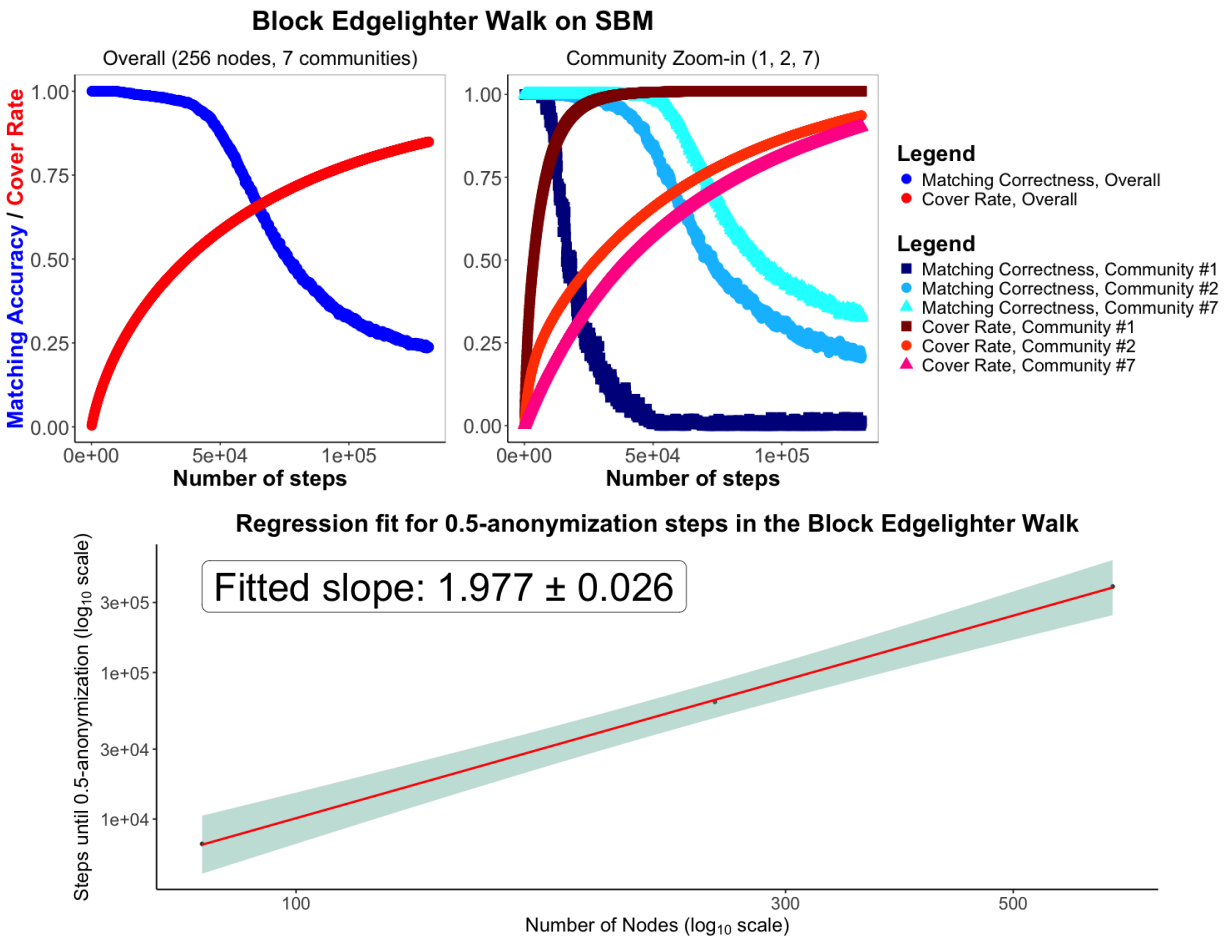

图 2：我们展示了在块边行者模型下，初始图服从随机块模型分布时，匹配正确性随 $n = 256$个顶点的步数变化情况。左上图显示整个网络的匹配正确性（蓝色）与步数的关系；右上图显示社区1（最小的社区）、社区2（从 $K - 2$ 个大小为 $n^{2/3}$的社区中随机选取的一个社区，此处为 $K = 7$）以及社区7（最大的社群）的匹配正确性与步数的关系。在所有这些图中，我们还绘制了边覆盖率（红色）随步数的变化。底部图中，我们绘制了达到0.5-匿名化所需步数与节点数的双对数图，并在这些点上叠加了拟合回归线。

为了在更具结构性的真实世界网络上进一步验证所提出的结果，我们在一个由大型欧洲研究机构生成的电子邮件通信网络上实施了分块边灯笼人游走。在该网络中，节点代表个人，边表示个人之间的电子邮件往来。每个人所属的部门均以数值部门标签作为节点属性提供，因此我们可以将每个部门视为一个独立的社区。有关该网络及其应用的更多详细信息，请参阅原始来源 [41]。我们从 https://snap.stanford.edu/data/email-Eu-core.html 获取了该网络。为简洁起见，并为了更好地与我们的理论

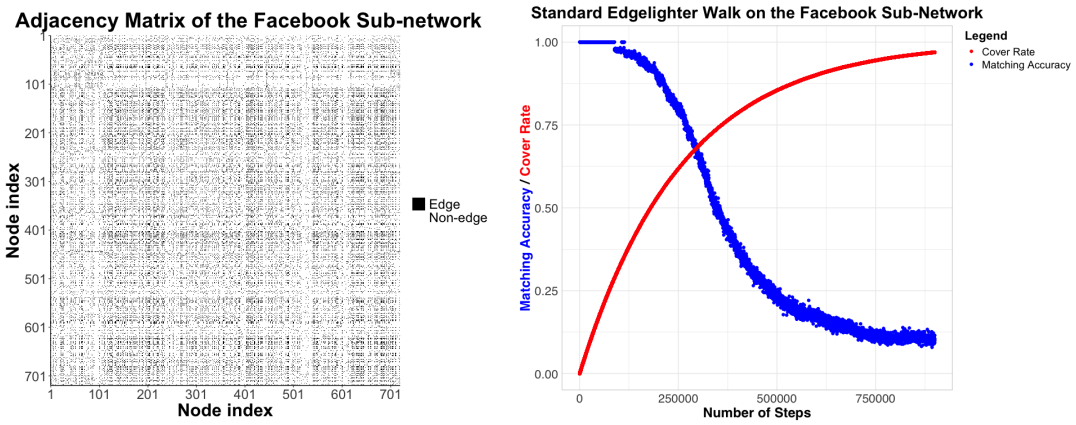

图 3：左：我们在其上执行 edgelighter 游走的 Facebook 子图的邻接矩阵。右：Facebook 子网络中，匹配正确性（蓝色）和覆盖率（红色）随步数变化的曲线。

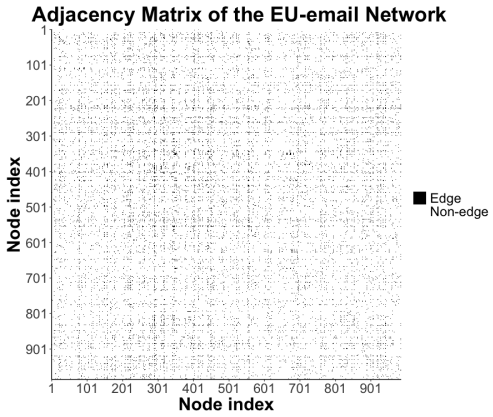

图 4：无向欧盟电子邮件通信网络的邻接矩阵。节点索引根据部门归属重新排序。

在该框架中，我们只考虑网络中最大的连通分量，并通过去除边的有向性将网络转化为无向图。所得邻接矩阵如图 4所示，其中我们根据社群对节点进行了重新排序。与前文讨论的随机块模型实验中的分块 edgelighter 游走一样，我们将定义 7 中的相同分块 edgelighter 游走应用于该网络；其中，我们将部门视为社群，并将每位电子邮件用户的部门归属作为其真实社群标签。为节省计算资源，每经过 220 步才与原始网络进行一次匹配。在图 5的左侧面板中，我们绘制了匹配正确性（蓝色）随步数变化的曲线，以及边覆盖率（红色）随步数变化的曲线。该图显示出与图 2中模拟随机块模型图上的分块 edgelighter 游走所观察到的类似模式。

为了更好地理解社区中的匿名化，图5的中间和右侧面板展示了匹配正确性随步数变化的图，以及由来自社区的节点诱导的子图的覆盖率随步数变化的曲线

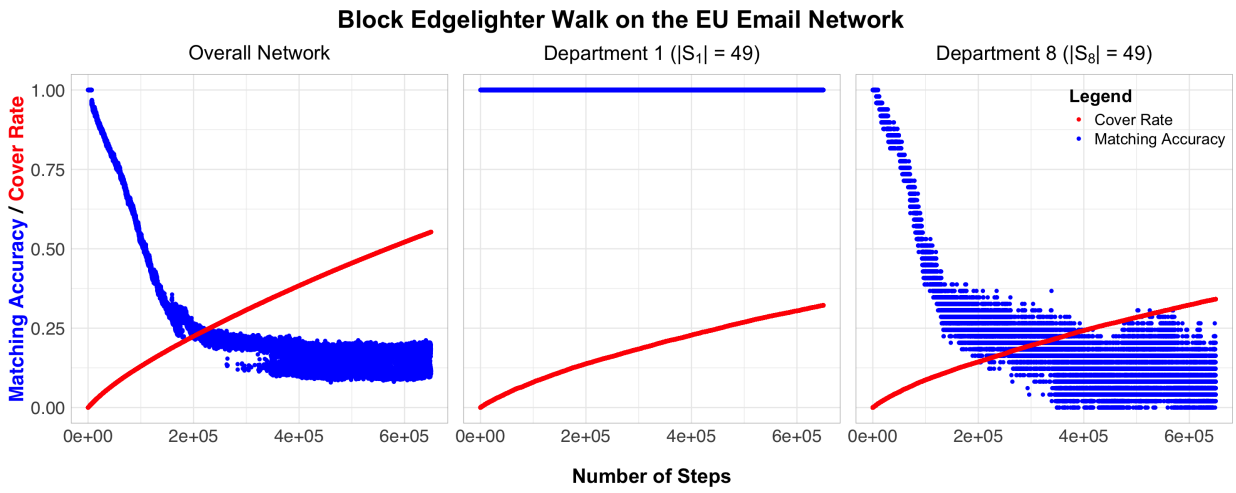

图 5：我们绘制了整个无向 EU 电子邮件网络（左图）以及由社区（部门）#1 的节点（中图）和社区（部门）#8 的节点（右图）诱导的子网络的匹配正确性与迭代图（蓝色），以及覆盖率与迭代曲线（红色）。

（部门）#1（中间面板，49 个成员）以及来自社区（部门）#8 的节点（右侧面板，49 个成员）诱导的子图。其他一些选定部门的图表见附录7.8.3。从图表中可以看出，某些社区中的匿名化发生得早得多。这一观察支持了我们的假设：结构化网络中更快的匿名化是局部现象，是某些社区匿名化所致；而我们认为，要在全局范围内破坏整个匹配，仍需要与噪声在全局范围内充分混合所需时间同一量级的时间。

6 结论与讨论

本文探讨了在不同图模型和不同灯夫式随机游走方案（我们称之为边灯人游走模型）下，图信号的匿名化时间与图上马尔可夫噪声混合时间之间的关系。通过考察理论界和仿真结果，我们发现，破坏原始图与其含噪对应图之间可匹配性所需的时间，与噪声的混合时间高度吻合，尤其是在 Erd˝os–R´enyi 随机图中。具体而言，结果表明，对于这些图上的标准边轻化游走，原始图与含噪图之间的匹配会在 $\Theta(n^2 \log n)$ 步之后减弱，这与边提灯人的混合时间一致。然而，在某些具有结构的图模型中，特别是社群结构不均衡的随机块模型（SBM）中，局部匿名化可能发生在噪声完全混合之前。这凸显了此类模型中截然不同的动态特性。我们的研究结果表明，在具有 n 个节点且社群规模差异足够大的随机块模型网络中，局部破坏可匹配性所需的时间约为某个 $\alpha < 2$ 对应的 $\Theta(n^\alpha \log n)$，显著早于稠密图的 $\Theta(n^2 \log n)$ 阈值。请注意，混合时间取决于图中的

边数。因此，对于较稀疏的图（边数少于 $\Theta(n^2)$），噪声的混合时间可能小于 $\Theta(n^2 \log n)$。这表明，社团结构的存在至少能加快局部匿名化过程，减少隐藏原始图结构所需的步数。本文旨在提供一个匿名化速度快于混合速度的简单示例；因此，我们将进一步探究这一背景下此类更稀疏图的工作留待未来开展。

为验证我们的理论预测，我们在合成网络和现实世界网络上开展了一系列模拟。模拟结果支持了我们的理论论断；回归分析表明，拟合得到的可匹配性曲线斜率与匿名化时间的预测主阶相符。我们进一步将分析扩展到现实世界网络，包括 Facebook 好友网络，以及来自一家欧洲研究机构的电子邮件通信网络。这些实验为所提出模型在真实数据上的表现提供了实践层面的洞见，证明了我们的理论结果不仅适用于合成环境。特别是，Facebook 网络上的标准 edgelighter 随机游走和电子邮件网络上的分块边灯笼人游走都表明，匿名化过程与预期的理论界一致，进一步彰显了我们理论的可靠性。

本研究的发现为未来研究开辟了新的方向。首先，尽管我们的结果表明，社区结构可以加快匿名化，但目前的结果仅基于边结构信息，因此值得探讨节点相似性在此情境下的影响。此外，将我们的模型扩展到其他类型的结构化图，例如分层随机块模型 [18, 47, 52, 58, 59], 具有重叠社区的块模型 [2], 或结构更复杂的图 [7, 35], ，或许能更深入地揭示图结构与匿名化时间之间的关系。此外，如果考虑时序网络集合，信号从一个图传播到下一个图时可能会产生依赖性，这超出了马尔可夫噪声的范畴。进一步分析存在时间依赖性时的可匹配性及其他统计推断任务，或能为基于网络的、更贴近现实数据的建模提供宝贵见解。最后，尽管种子图匹配（SGM）算法为将含噪图与其原图进行匹配提供了实用方案，但若能探索种子信息更先进的用途——不仅仅是将求解器限制在函数空间的某个特定子空间内——以提升匹配过程的效率和准确性，将会很有价值；参见 [56]相关方向的研究步骤。

参考文献

[1] Emmanuel Abbe。社区检测与随机块模型：近期进展。《机器学习研究杂志》，18(177):1–86，2018年。[2] E. M. Airoldi、D. M. Blei、S. E. Fienberg 和 E. P. Xing。混合成员随机块模型。《机器学习研究杂志》，9:1981–2014，2008年。[3] D. Aldous 和 J. A. Fill。可逆马尔可夫链与图上的随机游走，2002年。[4] MHD Samy Alnaimy 和 Mohammad Said Desouki。用于联合网络对齐与链接预测的扩展图嵌入。《大数据杂志》，9(1):41，2022年。[5] Nima Anari 和 Vijay V Vazirani。平面图完美匹配属于 NC。《*ACM*杂志》（*JACM*），67(4):1–34，2020年。[6] J. Arroyo、D. L. Sussman、C. E. Priebe 和 V. Lyzinski。含观测误差网络中的最大似然估计与图匹配。《计算统计学与图形统计学》，第1–13页，2021年。[7] A. Athreya、D. E. Fishkind、M. Tang、C. E. Priebe、Y. Park、J. T. Vogelstein、K. Levin、V. Lyzinski、Y. Qin 和 D. Sussman。随机点积图上的统计推断：综述。《机器学习研究杂志》，18:1–92，2018年。[8] Avanti Athreya、Zachary Lubberts、Youngser Park 和 Carey E Priebe。网络时间序列中的欧几里得镜像与动力学，2024年。[9] L. Babai。拟多项式时间内的图同构。载于第*48*届 *ACM* 计算理论年度研讨会论文集，第684–697页，2016年。[10] Sven Banisch 和 Sven Banisch。均匀混合下的投票者模型。《基于主体的模型的马尔可夫链聚合》，第57–82页，2016年。[11] Dave Bayer 和 Persi Diaconis。探寻燕尾洗牌的奥秘。《应用概率年刊》，第294–313页，1992年。[12] P. J. Bickel 和 A. Chen。网络模型以及 Newman–Girvan 等模块度的非参数观点。《美国国家科学院院刊》，106(50):21068–21073，2009年。[13] B. Bollob´as。随机图。Springer，1998年。[14] B. Bollob´as、S. Janson 和 O. Riordan。非齐次随机图中的相变。《随机结构与算法》，31:3–122，2007年。[15] Andries E Brouwer 和 Willem H Haemers。距离正则图。载于图的谱，第177–185页。Springer，2011年。

[16] Joshua Cape、Minh Tang 和 Carey E Priebe。《二到无穷范数与奇异子空间几何及其在高维统计中的应用》。《统计年鉴》，47(5)：2405–2439，2019年。[17] Fan Chung 和 Linyuan Lu。《集中不等式与鞅不等式：综述》。Internet mathematics，3(1)：79–127，2006年。[18] A. Clauset、C. Moore 和 M. E. J. Newman。《网络中的层次结构与缺失链接预测》。《自然》，453：98–101，2008年。[19]D. Conte、P. Foggia、C. Sansone 和 M. Vento。《模式识别中的图匹配：三十年回顾》。《国际模式识别与人工智能杂志》，18(03)：265–298，2004年。[20] D. Cullina 和 N. Kiyavash。《Erd˝os-R´enyi 图匹配的改进可达性界与反向界》。载于 ACM SIGMETRICS 性能评估评论，第44卷（1），第63–72页。ACM，2016年。[21]D. Cullina 和 N. Kiyavash。《相关 Erd˝os-R´enyi 图的精确对齐恢复》。arXiv 预印本 arXiv:1711.06783，2017年。[22] Riet De Smet 和 Kathleen Marchal。《当前网络推断方法的优势与局限》。《自然综述：微生物学》，8(10)：717–729，2010年。[23] Xinjie Du 和 Minh Tang。《随机图中潜在位置相等性的假设检验》。《伯努利》，29(4)：3221–3254，2023年。[24] P. Erd˝os 和 A. R´enyi。《论随机图，I》。《数学出版物》，6：290–297，1959年。[25] P. Erd˝os 和 A. R´enyi。《概率论中的一个经典问题》。《匈牙利科学院数学研究所学报》，6(1-2)：215–220，1961年。[26] Z. Fan、C. Mao、Y. Wu 和 J. Xu。《谱图匹配与正则化二次松弛：算法与理论》。载于国际机器学习会议，第2985–2995页。PMLR，2020年。[27] Zhou Fan、Cheng Mao、Yihong Wu 和 Jiaming Xu。《谱图匹配与正则化二次松弛：算法与理论》。载于国际机器学习会议，第2985–2995页。PMLR，2020年。[28] Farzad V Farahani、Waldemar Karwowski 和 Nichole R Lighthall。《应用图论识别人脑网络中的连接模式：系统综述》。《神经科学前沿》，13：585，2019年。[29] D.E. Fishkind、S. Adali、H.G. Patsolic、L. Meng、D. Singh、V. Lyzinski 和 C.E. Priebe。《种子图匹配》。《模式识别》，87：203–215，2019年。

[30] Bruna de Paula Fonseca e Fonseca、Ricardo Barros Sampaio、Marcus Vinicius de Ara´ujo Fonseca 和 Fabio Zicker。《健康研究中的合著网络分析：方法及潜在用途》。健康研究政策与系统，14：1–10，2016年。[31] A. Frieze 和 M. Karo´nski。《随机图导论》。剑桥大学出版社，2016年。[32] Matteo Frigo、Emilio Cruciani、David Coudert、Rachid Deriche、Emanuele Natale 和 Samuel Deslauriers-Gauthier。《网络对齐与相似性揭示结构连接组中基于图谱的拓扑差异》。网络神经科学，5(3)：711–733，2021年。[33] Cedric E Ginestet、Jun Li、Prakash Balachandran、Steven Rosenberg 和 Eric D Ko-laczyk。《功能性神经影像中的网络数据假设检验》。《应用统计年刊》，第725–750页，2017年。[34] Mark Heimann、Haoming Shen、Tara Safavi 和 Danai Koutra。《REGAL：基于表示学习的图对齐》。载于第*27*届*ACM*信息与知识管理国际会议论文集，第117–126页，2018年。[35] P. D. Hoff、A. E. Raftery 和 M. S. Handcock。《社会网络分析的潜在空间方法》。《美国统计学会杂志》，97(460)：1090–1098，2002年。[36] P. W. Holland、K. B. Laskey 和 S. Leinhardt。《随机块模型：初步探索》。《社会网络》，5(2)：109–137，1983年。[37] Saber Jafarizadeh。《最快混合可逆马尔可夫链：团提升图与子图》。《*IEEE*网络信号与信息处理汇刊》，6：88–104，2020年。[38] B. Karrer 和 M. E. J. Newman。《网络中的随机块模型与社团结构》。《物理评论*E*》，83，2011年。[39] Teresa L´azaro、Roger Guimer`a 和 Marta Sales-Pardo。《多个网络的概率对齐》。《自然通讯》，16(1)：3949，2025年4月。[40] Clement Lee 和 Darren J Wilkinson。《随机块模型及其图聚类扩展综述》。《应用网络科学》，4(1)：1–50，2019年。[41] Jure Leskovec、Jon Kleinberg 和 Christos Faloutsos。《图演化：密度增加与直径缩小》。《*ACM*数据知识发现汇刊》（*TKDD*），1(1)：2–es，2007年。[42] David A Levin、Malwina J Luczak 和 Yuval Peres。《均值场伊辛模型的Glauber动力学：截断现象、临界幂律与亚稳态》。《概率论及相关领域》，146：223–265，2010年。[43] David A Levin 和 Yuval Peres。《马尔可夫链与混合时间》，第107卷。美国数学学会，2017年。

[44] K. Levin、A. Avanti、M. Tang、V. Lyzinski 和 C. E. Priebe。多个随机点积图全体嵌入的中心极限定理。载于*2017*年 *IEEE* 国际数据挖掘会议研讨会（*ICDMW*），第 964–967 页。IEEE，2017年。[45] K. Levin、A. Lodhia 和 E. Levina。从具有未知边分布的多个网络中恢复共享结构。《机器学习研究杂志》，23(3)：1–48，2022年。[46] Guanpeng Li、Siva Kumar Sastry Hari、Michael Sullivan、Timothy Tsai、Karthik Pat- tabiraman、Joel Emer 和 Stephen W Keckler。理解深度学习神经网络（DNN）加速器及其应用中的误差传播。载于高性能计算、网络、存储与分析国际会议论文集，第 1–12 页，2017年。[47] Tianxi Li、Lihua Lei、Sharmodeep Bhattacharyya、Koen Van den Berge、Purnamrita Sarkar、Peter J Bickel 和 Elizaveta Levina。通过递归划分进行层次社区检测。《美国统计学会杂志》，117(538)：951–968，2022年。[48] L´aszl´o Lov´asz。图上的随机游走。组合数学，*Paul Erd˝os* 已八十岁，2(1-46)：4，1993年。[49] V. Lyzinski。通过图匹配从置乱图中恢复信息。*IEEE* 信息论汇刊，64(5)：3254–3273，2018年。[50] Vince Lyzinski、Sancar Adali、Joshua T Vogelstein、Youngser Park 和 Carey E Priebe。通过联合优化保真度与可通约性进行种子图匹配。统计学，1050：16，2014年。[51] Vince Lyzinski、Donniell E Fishkind、Marcelo Fiori、Joshua T Vogelstein、Carey E Priebe 和 Guillermo Sapiro。图匹配：自行放宽，风险自负。*IEEE* 模式分析与机器智能汇刊，38(1)：60–73，2015年。[52] Vince Lyzinski、Minh Tang、Avanti Athreya、Youngser Park 和 Carey E Priebe。层次随机块模型中的社区检测与分类。*IEEE* 网络科学与工程汇刊，4(1)：13–26，2016年。[53] J. J. McAuley 和 J. Leskovec。学习在自我中心网络中发现社交圈。载于*NIPS*，第 2012 卷，第 548–56 页，2012年。[54] Yuriy Mishchencko、Joshua T Vogelstein 和 Liam Paninski。利用钙荧光成像数据推断神经元连接性的贝叶斯方法。《应用统计年刊》，第 1229–1261 页，2011年。[55] Ben Morris、Weiyang Ning 和 Yuval Peres。卡片循环到随机洗牌的混合时间。*The Annals of Applied Probability*，24(5)：1835–1849，2014年。

[56] E. Mossel 和 J. Xu。通过大邻域统计进行种子图匹配。随机结构与算法，57(3)：570–611，2020年。[57] M. E. J. Newman。流行病在网络上的传播。物理评论 *E*，66：016128，2002年。[58] Majid Noroozi 和 Marianna Pensky。区块模型层级。*Sankhya A*，84(1)：64–107，2022年。[59] T. P. Peixoto。大型网络中的层次区块结构与高分辨率模型选择。物理评论 *X*，4，2014年。[60] C. E. Priebe、D. L. Sussman、M. Tang 和 J. T. Vogelstein。对含误差观测图进行统计推断。《计算与图形统计学杂志》，24(4)：930–953，2015年。[61] R. C. Read 和 D. G. Corneil。图同构疾病。图论杂志，1(4)：339–363，1977年。[62] Jes´us D Arroyo Reli´on、Daniel Kessler、Elizaveta Levina 和 Stephan F Taylor。网络分类及其在脑连接组学中的应用。《应用统计年刊》，13(3)：1648，2019年。[63] K. Riesen、X. Jiang 和 H. Bunke。精确与非精确图匹配：方法论与应用。图数据管理与挖掘，第 217–247 页，2010年。[64] K. Rohe、S. Chatterjee 和 B. Yu。谱聚类与高维随机块模型。统计年鉴，39：1878–1915，2011年。[65] Ayushi Saxena 和 Vince Lyzinski。迷失于洗牌之中：存在含误差网络顶点标签时的检验功效。计算统计学与数据分析，204：108091，2025年。[66] Frank R Schmidt、Eno Toppe 和 Daniel Cremers。高效平面图割及其在计算机视觉中的应用。载于*2009 年 IEEE 计算机视觉与模式识别会议*，第 351–356 页。IEEE，2009年。[67] F. Shirani、S. Garg 和 E. Erkip。种子图匹配：高效算法与理论保证。*arXiv* 预印本 *arXiv:1711.10360*，2017年。[68] Vishal Sood、Sidney Redner 和 Dani Ben-Avraham。Erd˝os–Rényi 随机图的首次通过性质。物理学杂志*A*：数学与通论，38(1)：109，2004年。[69] O. Sporns。发现人类连接组。麻省理工学院出版社，2012年。[70] Richard P. Stanley。枚举组合学。剑桥高等数学研究丛书。剑桥大学出版社，第 2 版，2011年。

[71] Ulrich Stelzl, Uwe Worm, Maciej Lalowski, Christian Haenig, Felix H Brembeck, Heike Goehler, Martin Stroedicke, Martina Zenkner, Anke Schoenherr, Susanne Koeppen 等。人类蛋白质相互作用网络：用于注释蛋白质组的资源。《细胞》，122(6):957–968，2005年。[72] Daniel L Sussman、Minh Tang、Donniell E Fishkind 和 Carey E Priebe。随机区块模型图的一致邻接谱嵌入。《美国统计学会杂志》，107(499):1119–1128，2012年。[73] Reza Takapoui 和 Stephen Boyd。图同构问题的线性规划启发式方法。*arXiv* 预印本 *arXiv:1611.00711*，2016年。[74] M. Tang、A. Athreya、D. L. Sussman、V. Lyzinski、Y. Park 和 C. E. Priebe。随机图的半参数双样本假设检验问题。《计算统计学与图形统计学杂志》，26(2):344–354，2017年。[75] M. Tang、D. L. Sussman 和 C. E. Priebe。潜在位置图的普遍一致顶点分类。《统计年鉴》，41(3):1406–1430，2013年。[76] S. Umeyama。加权图匹配问题的特征分解方法。*IEEE* 模式分析与机器智能汇刊，10(5):695–703，1988年。[77] Fabrizio Vecchio、Francesca Miraglia 和 Paolo Maria Rossini。连接组：图论在功能性脑网络架构中的应用。《临床神经生理学实践》，2:206–213，2017年。[78] J. T. Vogelstein 和 C. E. Priebe。乱序图分类：理论与连接组应用。《分类杂志》，32(1):3–20，2015年。[79] Joshua T Vogelstein、John M Conroy、Vince Lyzinski、Louis J Podrazik、Steven G Kratzer、Eric T Harley、Donniell E Fishkind、R Jacob Vogelstein 和 Carey E Priebe。用于图匹配的快速近似二次规划。《公共科学图书馆·综合》，10(4):e0121002，2015年。[80] J.T. Vogelstein、J.M. Conroy、V. Lyzinski、L.J. Podrazik、S.G. Kratzer、E.T. Harley、D.E. Fishkind、R.J. Vogelstein 和 C.E. Priebe。用于图匹配的快速近似二次规划。《公共科学图书馆·综合》，10(04)，2014年。[81] Jinhui Wang、Xinian Zuo、Zhengjia Dai、Mingrui Xia、Zhilian Zhao、Xiaoling Zhao、Jian-ping Jia、Ying Han 和 Yong He。阿尔茨海默病风险人群的功能性脑连接组受损。《生物精神病学》，73(5):472–481，2013年。[82] Jinhui Wang、Xinian Zuo 和 Yong He。静息态功能磁共振成像的基于图的网络分析。《系统神经科学前沿》，4:1419，2010年。[83] S. Wang、J. Arroyo、J. T. Vogelstein 和 C. E. Priebe。图的联合嵌入。*IEEE* 模式分析与机器智能汇刊，43(4):1324–1336，2019年。

[84] David Bruce Wilson。Rudvalis 洗牌的混合时间。*Electronic Communicationsin Probability*，8:77–85，2003年。[85] Y. Wu、J. Xu 和 S. H. Yu。确定随机图匹配的精确重构阈值。*IEEE* 信息论汇刊，68(8)，2022年。[86] Yihong Wu、Jiaming Xu 和 Sophie H Yu。无标记随机图的相关性检验。*The Annals of Applied Probability*，33(4):2519–2558，2023年。[87] J. Yan、X. Yin、W. Lin、C. Deng、H. Zha 和 X. Yang。图匹配近期进展简述。载于*2016*年 *ACM* 多媒体检索国际会议论文集，第167–174页。ACM，2016年。[88] Lyudmila Yartseva 和 Matthias Grossglauser。渗流图匹配的性能。载于首届 *ACM* 在线社交网络会议论文集，第119–130页，2013年。[89] Si Zhang 和 Hanghang Tong。FINAL：快速属性网络对齐。载于第*22*届 *ACM SIGKDD* 国际知识发现与数据挖掘会议论文集，第1345–1354页，2016年。[90] Zhen Zhang、Yijian Xiang、Lingfei Wu、Bing Xue 和 Arye Nehorai。KERGM：核化图匹配。神经信息处理系统进展，32，2019年。[91] Xiaohan Zou。知识图谱应用综述。载于*Journal of Physics:Conference Series*，第1487卷，文章编号012016。IOP Publishing，2020年。

7 附录

本文在此汇集论文主要结果的证明。

7.1 引理的证明1

首先注意，标准 Edgelighter 马尔可夫链的平稳分布的归一化常数 π° 为

$$\Psi = \sum_{(x,c)} \frac{1}{n} q_2^{\#\text{ of on edges in } c} q_1^{\#\text{ of off edges in } c} = \begin{cases} \frac{q_2^{\binom{n}{2}+1} - q_1^{\binom{n}{2}+1}}{q_2 - q_1} & \text{if } q_2 \neq q_1; \\ (\binom{n}{2}+1) q_2^{\binom{n}{2}} & \text{if } q_2 = q_1. \end{cases}$$

无论哪种情况，该归一化常数都不大于 1。我们将紧密遵循 [43, 第 19.3] 节中定理 19.2 的证明；为使论述完整，现提供必要的细节。对于上界，令 $\vec{0}$ 为所有边均关闭的构型，$\vec{1}$ 为所有边均开启的构型。令 y 使 $\mathbf{P}_x(L_t = \cdot \,|\tau_{(u)\text{覆盖}} \leq t)$ 最小（因此 $\mathbf{P}_x(L_t = y|\tau_{\text{cover}}^{(u)} \leq t) \leq 1/n$），并注意

$$\begin{aligned} \frac{\mathbb{P}_{(x,\vec{0})}(Z_t = (y,\vec{1}))}{\pi^\circ(y,\vec{1})} &= \frac{\mathbb{P}_x(\tau_{\text{cover}}^{(u)} \leq t) q_2^{\binom{n}{2}} \mathbb{P}_x(L_t = y|\tau_{\text{cover}}^{(u)} \leq t)}{\frac{1}{n} q_2^{\binom{n}{2}} / \Psi} \\ &\leq n\Psi \mathbb{P}_x(\tau_{\text{cover}}^{(u)} \leq t) \frac{1}{n} \\ &\leq \mathbb{P}_x(\tau_{\text{cover}}^{(u)} \leq t) \end{aligned}$$

由此可得，(Z_t) 的分离距离满足

$$s^\circ(t) = \max_{(x,c),(y,c')} 1 - \frac{\mathbb{P}_{(x,c)}(Z_t = (y,c'))}{\pi^\circ((y,c'))} \geq \mathbb{P}_x(\tau_{\text{cover}}^{(u)} > t).$$

这里，我们还可以得到，如果

$$\begin{aligned} d^\circ(t) &= \max_{(x,c)} \|\mathbb{P}_{(x,c)}^t(\cdot) - \pi^\circ\|_{TV} \\ \bar{d}^\circ(t) &= \max_{(x,c),(y,c')} \|\mathbb{P}_{(x,c)}^t(\cdot) - \mathbb{P}_{(y,c')}^t(\cdot)\|_{TV} \end{aligned}$$

那么，由于 (Z_t) 是可逆的，

$$\begin{aligned} d^\circ(t) \leq \bar{d}^\circ(t) \leq 2d^\circ(t) \\ s^\circ(2t) \leq 1 - (1 - \bar{d}^\circ(t))^2 \end{aligned}$$

¯其中最后一行的不等式由 [3]中的引理 4.7 得出。因此，由于 $d^\circ(t)$具有次乘性（见 [43]中的引理 4.11），并且注意到根据定义 5，$d^\circ(t_m^\circ) \leq 1/4$，

$$\bar{d}^\circ(2t_m^\circ) \leq (\bar{d}^\circ(t_m^\circ))^2 \leq 4d^\circ(t_m^\circ)^2 \leq 1/4$$

因此，对于任意 x,

$$\mathbb{P}_x(\tau^{(u)}_{\text{cover}} > 4t^\circ_m) \le s^\circ(4t^\circ_m) \le 1 - (1 - \bar{d}^\circ(2t^\circ_m))^2 < 1/2.$$

在每个 $4t^\circ_m$ 的整数倍处依次应用马尔可夫性质，即可得到对所有整数 $k \ge 1$,

$$\mathbb{P}_x(\tau^{(u)}_{\text{cover}} > 4kt^\circ_m) \le \mathbb{P}_x(\tau^{(u)}_{\text{cover}} > 4t^\circ_m)^k < (1/2)^k.$$

因此

$$\mathbb{E}_x(\tau^{(u)}_{\text{cover}}) = \sum_{k=0} \mathbb{P}_x(\tau^{(u)}_{\text{cover}} > k) \le 4t^\circ_m \sum_{k=0} \mathbb{P}_x(\tau^{(u)}_{\text{cover}} > k4t^\circ_m) \le 8t^\circ_m.$$

对于下界，我们再次紧密沿用 [43, 第 19 节3] 中的论证。令 $Z_t = (L_t, C_t)$ 表示边提灯人在时间 t 的状态。设 $w = 8t^{(u)}_{\text{cover}} + t_m(L_t, 1/8)$（其中 π_L 是 (L_t) 的平稳分布——即 V 上的均匀概率，而 $t_m(L_t, 1/8)$ 是满足 $\max_x \|\mathbb{P}_x(L_t \in \cdot) - \pi_L\|_{TV} \le 1/8$ 的最小时间）。固定 $(v, \vec{0})$ 并定义

$$\mu_s = \mathbb{P}_{(v,\vec{0})}(Z_w \in \cdot | \tau^{(u)}_{\text{cover}} = s)$$

于是我们有

$$\mathbb{P}_{(v,\vec{0})}(Z_w \in \cdot) = \sum_s \mu_s \mathbb{P}_{(v,\vec{0})}(\tau^{(u)}_{\text{cover}} = s),$$

因此

$$\|\mathbb{P}_{(v,\vec{0})}(Z_w \in \cdot) - \pi^\circ\|_{TV} \le \sum_s \mathbb{P}_{(v,\vec{0})}(\tau^{(u)}_{\text{cover}} = s)\|\mu_s - \pi^\circ\|_{TV}$$

由马尔可夫不等式可得 $\mathbb{P}(\tau^{(u)}_{\text{cover}} > 8t^{(u)}_{\text{cover}}) \le 1/8$，从而得到

$$\|\mathbb{P}_{(v,\vec{0})}(Z_w \in \cdot) - \pi^\circ\|_{TV} \le 1/8 + \sum_{s \le 8t^{(u)}_{\text{cover}}} \mathbb{P}_{(v,\vec{0})}(\tau^{(u)}_{\text{cover}} = s)\|\mu_s - \pi^\circ\|_{TV}$$

如果 $\tau^{(u)}_{\text{cover}} = s$ 对于 $s \le w$，那么 C_w 具有如下分布（对于一个构型 $c \in C$，令 m_c 和 r_c 分别表示 c中 1 和 -1的数量）

$$\mathbb{P}(C_w = \vec{c}) = \nu(c) \propto q_2^{m_c} q_1^{r_c}$$

与 L_w的位置无关。因此

$$\begin{aligned}\mu_s &= \sum_x \mathbb{P}_{(v,\vec{0})}(Z_w \in \cdot | \tau^{(u)}_{\text{cover}} = s, L_s = x)\mathbb{P}_{(v,\vec{0})}(L_s = x | \tau^{(u)}_{\text{cover}} = s) \\ &= \sum_x (\mathbb{P}_x(L_{w-s} \in \cdot) \times \nu)\mathbb{P}_{(v,\vec{0})}(L_s = x | \tau^{(u)}_{\text{cover}} = s)\end{aligned}$$

由于 $\pi^\circ = \pi_L \times \nu$，我们有

$$\begin{aligned}\|\mu_s - \pi^\circ\|_{TV} &\le \sum_x \|\mathbb{P}_x(L_{w-s} \in \cdot) \times \nu - \pi^\circ\|_{TV} \mathbb{P}_{(v,\vec{0})}(L_s = x | \tau^{(u)}_{\text{cover}} = s) \\ &\le \max_x \|\mathbb{P}_x(L_{w-s} \in \cdot) - \pi_L\|_{TV}\end{aligned}$$

对于 $s \le 8t^{(u)}_{\text{cover}}$, $w - s > t_m(L_t, 1/8)$，因此 $\|\mu_s - \pi^\circ\|_{TV} \le 1/8$。因此，

$$\|\mathbb{P}_{(v,\vec{0})}(Z_w \in \cdot) - \pi^\circ\|_{TV} \le 1/4,$$

和 $w \ge t^\circ_m$。由 d 的次乘性可得

$$t_m(L_t, 1/8) \le 3t_m(L_t) \le 3 \le 3t^{(u)}_{\text{cover}},$$

其中，第二个不等式由 [3], 第4章以及 T_j 是状态 j的命中时间这一事实推出。因此，

$$t^\circ_m \le w \le 11t^{(u)}_{\text{cover}}$$

正如所需。

7.2 方程(5)和(6)的推导

7.2.1 证明方程(6)

我们旨在证明关于$\mathfrak{p}_{t,\ u,\ v}$的给定界。首先，我们仿照标准边轻化过程定义全局边缘照亮游走，不同之处在于边提灯人的位置（即 L_t）无关紧要。也就是说，在第（$t+1$）步，全局边照明器会随机选择任意一个 $e \in \binom{V}{2}$，并按照方程（$h_{G_{t+1}}(e)$ 的定义来定义2）和（3。考虑给定 $\{u, v\} \in \binom{V}{2}$时的三个事件:

$\mathcal{E}_{1,\ t} = \{$标准行走中的边提灯人截至时间 $t\}$ 尚未经过 $\{u, v\}$ $\mathcal{E}_{2,\ t} = \Big\{$标准行走中的边提灯人在 $(\{L_{2i}, L_{2i+1}\})$ 中尚未经过 $\{u, v\}_{i=0}^{\lfloor \frac{t-1}{2} \rfloor}\Big\}$$\mathcal{E}_{3,\ t} = \{$全局行走中的边提灯人截至 $t\}$ 尚未经过 $\{u, v\}$

于是我们得到

$$\left(1 - \frac{1}{\binom{n}{2}}\right)^t = \mathbb{P}(\mathcal{E}_{3,t}) \le \mathbb{P}(\mathcal{E}_{1,t}) \le \mathbb{P}(\mathcal{E}_{2,t}) = \left(1 - \frac{1}{\binom{n}{2} + n}\right)^{\lfloor \frac{t+1}{2} \rfloor}$$

利用对于 $n > 3$,

$$\left(1 - \frac{1}{\binom{n}{2} + n}\right) \le e^{-\frac{1}{\binom{n}{2}+n}} \quad \text{and} \quad \left(1 - \frac{1}{\binom{n}{2}}\right) \ge e^{-\frac{1}{\binom{n}{2}-1}},$$

我们有

$$\mathfrak{p}_{t,u,v} \le \exp\left\{-\frac{\lfloor \frac{t+1}{2} \rfloor}{\binom{n}{2} + n}\right\} \quad \text{and} \quad \mathfrak{p}_{t,u,v} \ge \exp\left\{-\frac{t}{\binom{n}{2} - 1}\right\},$$

这建立了方程（6）。

7.2.2 证明方程。(5)

令 $\tau_{(g)\text{cover}}$表示全局边照明器的覆盖时间，令 τ_{cover}表示方程4中定义的覆盖时间，则有 $\tau_{(g)\text{cover}} \leq_{st.} \tau^{(u)}_{\text{cover}} \leq_{st.} 2\tau_{\text{cover}}$（其中 $\leq_{st.}$表示随机序）。因此，对于 n充分大的情况，由简单的优惠券收集者渐近可得，

$$\frac{1}{2}n^2 \log n \leq \mathbb{E}\tau^{(g)}_{\text{cover}} \leq \max_x \mathbb{E}_x \tau^{(u)}_{\text{cover}} \leq \max_x \mathbb{E}_x \tau_{\text{cover}} \leq \frac{5}{2}n^2 \log n,$$

由此得到方程（5）。

7.3 定理1的证明

这里，我们将遵循标准的集中不等式理论来建立可匹配性结果。设 $B = A_t$ 和 $A = A_0$，对于一个固定的置换矩阵 P及其对应的置换 σ_P，考虑以下量

$$\begin{aligned} S_{n,P} &:= \operatorname{Tr}\left(APBP^T\right) - \operatorname{Tr}(AB) = \operatorname{Tr}\left(A[PBP^T - B]\right) \\ &= 2 \sum_{\substack{\{i,j\} \in \binom{V}{2} \text{ s.t.} \\ \{i,j\} \neq \{\sigma_P(i), \sigma_P(j)\}}} A_{i,j} \left(B_{\sigma_P(i),\sigma_P(j)} - B_{i,j}\right). \end{aligned}$$

直接计算可得

$$\begin{aligned} \mathbb{E}S_{n,P} &= 2 \sum_{\substack{\{i,j\} \in \binom{V}{2} \text{ s.t.} \\ \{i,j\} \neq \{\sigma_P(i), \sigma_P(j)\}}} \mathbb{E}\left\{\left(B_{\sigma_P(i),\sigma_P(j)} - B_{i,j}\right) \mid A_{i,j} = 1\right\} \mathbb{P}(A_{i,j} = 1) \\ &= 2p \sum_{\substack{\{i,j\} \in \binom{V}{2} \text{ s.t.} \\ \{i,j\} \neq \{\sigma_P(i), \sigma_P(j)\}}} \left\{\mathbb{E}\left(B_{\sigma_P(i),\sigma_P(j)} \mid A_{i,j} = 1\right) - \mathbb{E}\left(B_{i,j} \mid A_{i,j} = 1\right)\right\} \end{aligned}$$

注意到 $\mathfrak{p}_{t,\,i,\,j}$ 在这种情况下不依赖于 i，j，因此我们省略这些下标，后文记作 $\mathfrak{p}_t$。现在，我们通过对经过 t 步后指标 i和 j 的不同情况进行分析，来计算各个条件期望。

1. 我们首先考虑 $\mathrm{E}\left\{B_{\sigma_P(i)\ \sigma_P(j)} \mid A_{i,\,j} = 1\right\}$。如果 $\{\sigma_P(i)\ \sigma_P(j)\}$ 在时间 t 前已被边较轻的遍历，那么无论 $A_{\{\sigma_P(i)\ \sigma_P(j)\}}$ 和 $A_{i,\,j}$ 的状态如何，该边出现的概率都是 p。因此，这个条件期望等于

$$\mathbb{E}\left\{B_{\sigma_P(i),\sigma_P(j)} \mid A_{i,j} = 1\right\} = p(1 - \mathfrak{p}_t) + p\mathfrak{p}_t = p$$

2. 接下来考虑 $\mathrm{E}\{B_{i,j} \mid A_{i,j} = 1\}$。如果边提灯人在时间$t$之前已经遍历了 $\{i, j\}$，那么无论A_i的状态如何，该边的概率都是 $p_{\,,\,j}$；否则$B_{i,\,j} = A_{i,\,j}$。因此，这个条件期望等于

$$\mathbb{E}\left\{B_{i,j} \mid A_{i,j} = 1\right\} = p(1 - \mathfrak{p}_t) + \mathfrak{p}_t$$

综合以上结果，我们得到

$$\mathbb{E}(S_{n,P}) = 2\underbrace{\left|\left\{\{i,j\} \in \binom{V}{2} \text{ s.t. } \{i,j\} \neq \{\sigma_P(i), \sigma_P(j)\}\right\}\right|}_{=:\eta_P} \mathfrak{p}_t p(p-1)$$

如果 P打乱了 k 个顶点，那么 $\eta_P \in [(n-2)k/2,\ nk]$ 而这个期望可以界定为

$$(n-2)k\mathfrak{p}_t p(p-1) \leq \mathbb{E}(S_{n,P}) \leq 2nk\mathfrak{p}_t p(p-1).$$

More关于 $S_{n,\ P}$，可以将其表示为以下随机变量的函数 :

i. 求和中涉及的至多 $2nk$ 条边，$A_{i,\ j}$。改变其中任意一条，最多会使求和结果 $S_{n,\ P}$ 改变 4。

ii. 至多 $2(t+1)$ 个随机变量，它们决定了边提灯人行走的步数和边翻转。改变边提灯人的单步移动，至多会改变 $B_{i,\ j}$ 中的 2 个，因此被加数最多会改变 8。改变边提灯人的一次翻转，至多会改变一个 $B_{i,\ j}$，因此被加数最多会改变 2。

根据 McDiarmid不等式，对于 n 足够大时（回想一下，我们假设存在一个常数 $c > 0$，使得 $t \leq cn^2 \log n$）

$$\begin{aligned}
\mathbb{P}(S_{n,P} \geq 0) &\leq \mathbb{P}(|S_{n,P} - \mathbb{E}(S_{n,P})| \geq |\mathbb{E}S_{n,P}|) \\
&\leq 2\exp\left\{-\frac{(n-2)^2k^2p^2(p-1)^2\exp\left\{-\frac{4t}{n(n-1)-2}\right\}}{16(nk+t+1)}\right\} \\
&\leq 2\exp\left\{-\frac{(n-2)^2k^2p^2(p-1)^2n^{-5c}}{16(nk+cn^2\log n+1)}\right\}.
\end{aligned}$$

令 $\Pi_{n,\ k}$ 为恰好打乱 k 个顶点标签的置换集合，并定义事件

$$\beta_{k,t} = \{\exists P \in \Pi_{n,k} \text{ s.t. } S_{n,P} \geq 0\}.$$

对 $\Pi_{n,\ k}$ 中至多 n^k 个置换矩阵应用并合界，可得

$$\mathbb{P}(\beta_{k,t}) \leq 2\exp\left\{-\frac{(n-2)^2k^2p^2(p-1)^2n^{-5c}}{16(nk+cn^2\log n+1)} + k\log n\right\}.$$

回顾一下，根据假设，存在一个常数 $\alpha > 5c$ 使得 $k \geq n^\alpha$，我们有

$$\frac{1}{k\log n}\ \frac{(n-2)^2k^2p^2(p-1)^2n^{-5c}}{16(nk+cn^2\log n+1)} = \Omega\left(\frac{p^2(p-1)^2n^{\alpha-5c}}{\log^2 n}\right) = \omega(1).$$

因此，对此类 k 取并集可得

$$\mathbb{P}\left(\bigcup_{k\geq n^\alpha}\beta_{k,t}\right) \leq 2\exp\left\{-\Omega\left(\frac{n^{2\alpha-5c}}{\log n}\right) + \log n\right\} \leq 2\exp\left\{-\Omega\left(\frac{n^{2\alpha-5c}}{\log n}\right)\right\},$$

完成证明。

7.4 定理2的证明

稍微滥用一下记号，我们用 Π_n 同时表示 $\{1,2,\cdots,n\}$ 的置换集合（用小写希腊字母表示）和 $n\times n$ 置换矩阵集合（用大写罗马字母表示），具体含义由上下文决定。在不失一般性的前提下，我们在整个证明中假设 n^β 是一个整数；否则，只需在接下来的论证中将 n^β 全部替换为 $\lfloor n^\beta\rfloor$。若置换 $\sigma,\tau\in\Pi_n$ 在位置 $\ell\in\{1,2,\cdots,n\}$ 处不一致，则称它们在该位置不一致，即 $\sigma(\ell)\neq\tau(\ell)$。考虑 $\mathfrak{n}_\beta=\lfloor n/(2n^\beta+1)\rfloor$ 个两两不相交的集合，它们都属于 $\{1,2,\cdots,n\}$，每个集合的大小为 $2n^\beta+1$（最后一个集合可能多包含几个元素）。我们用 $\{S_i\}_{i=1}^{\mathfrak{n}_\beta}$ 表示这些集合。对每个 $i\in\{1,2,\cdots,\mathfrak{n}_\beta\}$，令 $\sigma^{(i)}$ 为 $\{1,2,\cdots,n\}$ 的一个置换，它固定 $\cup_{j:j\neq i}S_j$ 的所有元素，并且是一个循环错排（按循环分解来看，也就是 S_i 中长度为 $2n^\beta+1$ 的一个循环；但最后一个循环可能更长）。

注意，若 $(\sigma^{(i)})^k$ 表示置换

$$(\sigma^{(i)})^k=\underbrace{\sigma^{(i)}\circ\cdots\circ\sigma^{(i)}}_{k\text{ compositions}},$$

那么对于所有满足 $k_1\neq k_2$ 的 $k_1,k_2\in\{1,2,\cdots,2n^\beta\}$，$(\sigma^{(i)})^{k_1}$ 与 $(\sigma^{(i)})^{k_2}$ 至少在 $2n^\beta+1$ 个位置上不同（并且当所有 $i\neq\mathfrak{n}_\beta$ 时，恰好在 $2n^\beta+1$ 个位置上不同）。考虑所有形如

$$\sigma=(\sigma^{(1)})^{k_1}\circ(\sigma^{(2)})^{k_2}\circ\cdots\circ(\sigma^{(\mathfrak{n}_\beta)})^{k_{\mathfrak{n}_\beta}}\tag{10}$$

对于 $(k_1,k_2,\cdots,k_{\mathfrak{n}_\beta})\in\{1,2,\cdots,2n^\beta\}^{k_{\mathfrak{n}_\beta}}$。任意一对此类 σ，若其 $(k_1,k_2,\cdots,k_{\mathfrak{n}_\beta})$ 序列不同（即存在至少一个 j，使得它们的第 j 项不同），则它们至少在 $2n^\beta+1$ 个位置上不同。我们用 $\widetilde{\Pi}^\beta$ 表示式（10）中满足 $(k_1,k_2,\cdots,k_{\mathfrak{n}_\beta})\in\{1,2,\cdots,2n^\beta\}^{k_{\mathfrak{n}_\beta}}$ 的所有置换构成的集合。注意，$\widetilde{\Pi}^\beta$ 的大小为

$$\left|\widetilde{\Pi}^\beta\right|=(2n^\beta)^{\mathfrak{n}_\beta}\geq(2n^\beta)^{n/(2n^\beta+1)-1}.\tag{11}$$

对于一个置换集合置换 $\mathcal{Q}\subset\Pi_n$，将置换 $P\circ\mathcal{Q}$ 的作用定义为该序列 t

$$P\circ\mathcal{Q}=\{Q\in\Pi_n\text{ s.t. }P^{-1}Q\in\mathcal{Q}\}.$$

对于置换 $P\in\Pi_n$，定义事件

$$\mathcal{E}_{\beta,P}=\left\{\arg\max_{Q\in\Pi_n}\mathrm{Tr}\left(A_0QA_tQ^T\right)\subset P\circ\left(\bigcup_{k=0}^{n^\beta}\Pi_{n,k}\right)\right\}.$$

注意到，集合 $\{\mathcal{E}_{\beta,P}\}_{P\in\widetilde{\Pi}^\beta\cup\{I_n\}}$ 两两不相交，正如对于 $P\in\widetilde{\Pi}^\beta\cup\{I_n\}$，集合 $P\circ\left(\bigcup_{k=0}^{n^\beta}\Pi_{n,k}\right)$ 也两两不相交。为说明这一点，反设对于不同的 $P_1,P_2\in\widetilde{\Pi}^\beta\cup\{I_n\}$，存在一个 $Q\in P_1\circ\left(\bigcup_{k=0}^{n^\beta}\Pi_{n,k}\right)\cap P_2\circ\left(\bigcup_{k=0}^{n^\beta}\Pi_{n,k}\right)$。那么存在 $Q_1,Q_2\in\bigcup_{k=0}^{n^\beta}\Pi_{n,k}$，使得 $P_1^{-1}Q=Q_1$ 且 $P_2^{-1}Q=Q_2$，因此 $P_1=P_2Q_2Q_1^{-1}$。这

就意味着 P_1 和 P_2 至多会在 $2n^\beta$ 个位置上不同，但这与我们的假设矛盾： $P_1, P_2 \in \widetilde{\Pi}^\beta \cup \{I_n\}$ 是不同的，因为根据构造， $\widetilde{\Pi}^\beta \cup \{I_n\}$ 中不同的元素至少会在 $2n^\beta + 1$ 个位置上不同。

回顾一下，在第 3 节开头，我们提到，对于 $t \geq \tau_{\text{cover}}^{(u)}$， $A_t \sim \text{ER}(n,\ p)$ 与 A_0 相互独立，因此 $\mathbf{Tr}\left(A_0 P A_t P^T\right)$ 与任意置换 $A_0 A_t$) 的分布相同。 $P \in \Pi_n$ 由此可知， $\mathbb{P}(\mathcal{E}_{\beta,}$ $_P \mid t > \tau_{\text{cover}}^{(u)})$ 对所有 $P \in \Pi_n$ 都相同。因此，根据公式（11），

$$\mathbb{P}(\mathcal{E}_{\beta,I_n} \mid\ t > \tau_{\text{cover}}^{(u)}) \leq (2n^\beta)^{1-n/(2n^\beta+1)}. \tag{12}$$

定理 2 可通过考虑以下情况得证：

$$\begin{aligned}\mathbb{P}(\mathcal{E}_{\beta,I_n}) =& \mathbb{P}(\mathcal{E}_{\beta,I_n} | \tau_{\text{cover}}^{(u)} \geq t)\mathbb{P}(\tau_{\text{cover}}^{(u)} \geq t) + \mathbb{P}(\mathcal{E}_{\beta,I_n} | t > \tau_{\text{cover}}^{(u)})\mathbb{P}(t > \tau_{\text{cover}}^{(u)}) \\ \leq& \mathbb{P}(\tau_{\text{cover}}^{(u)} \geq t) + \mathbb{P}(\mathcal{E}_{\beta,I_n} \mid\ t > \tau_{\text{cover}}^{(u)})\end{aligned}$$

并结合公式（12），以及公式（7）中的界，利用假设 $t \geq 4n^2 \log n$。

7.5 定理 3 的证明3

对于每个 $i \in [K_n]$，令 n_i 表示第 i 个社区的大小。我们将使用以下切尔诺夫型界，该界改编自 [17]）中的定理3.2：设 $X_i \sim$ 服从伯努利分布(p_i)，且为相互独立的随机变量，满足 $X = \sum_i X_i$，并且 $\mathbb{E}(X) = \sum_i p_i$。那么，对于任意 $t > 0$,

$$\mathbb{P}\left(|X - \mathbb{E}(X)| > t\right) \leq 2\exp\left\{-\frac{t^2}{2\mathbb{E}(X) + 2t/3}\right\}.$$

将此应用于定义为社区 i 与社区 j 之间边数的 $m_{ij} \sim \mathbf{Bin}(n_i n_j, \Lambda_{i,j})$，并对 $m_{ii} \sim \mathbf{Bin}(\binom{n_i}{2}, \Lambda_{i,i})$ 作类似定义，我们有：以至少 $1 - K_n^2 n^{-3}$ 的概率（根据假设，$n_i n_j \Lambda_{i,j} \geq 5\log\ n$），

$$|m_{ij} - n_i n_j \Lambda_{i,j}| \leq \sqrt{5 n_i n_j \Lambda_{i,j} \log n} \ \text{ and} \tag{13}$$

$$\left| m_{ii} - \binom{n_i}{2}\Lambda_{i,i} \right| \leq \sqrt{5\binom{n_i}{2}\Lambda_{i,i} \log n}. \tag{14}$$

在证明的其余部分，我们将以方程（13）和（14）中的事件为条件，并将该事件记为 $\mathcal{A}_n$，因此社区1与社区 j 之间有$m_{1,j} \in n_1 n_j \Lambda_{1,j} \pm \sqrt{5 n_1 n_j \Lambda_{1,j} \log\ n}$ 条边。回顾一下，$\Lambda_{1,j} \leq b_3 n^{a_2}/n_1 n_j$，因此如果上述事件成立，就存在一个常数 $b_3' > 0$，使得 $m^* = \max_j m_{1,j} \leq b_3' n^{a_2}$。

考虑事件 $\mathcal{A}_n$ 上随机图的边提灯人游走。如果行走者位于社区1，那么在下一步中，边提灯人以概率 $1/2$ 移动到除社区1以外的一个均匀随机社区。记离开社区1之前的步数为 T_1，注意到

$$\mathbb{P}\left(T_1 > K_n m^*\right) = \left(\frac{1}{2}\right)^{K_n m^*} \leq \exp\{-m^*\},$$

因此，按照定理中的定义， T_1 以高概率是低于 $N = n^{1+a_2}\log^{3/2} n$ 的阶。因此， T_1 对整体混合时间影响甚微，我们可以不失一般性地假设边提灯人从社区1以外的位置出发。

边提灯人首次离开社区1后（若边提灯人开始时不在社区1内，则将此时刻视为时间0），我们定义如下更新过程：如果边提灯人进入社区1，立即移动到社区1内的另一个顶点，然后最终离开社区1，就称发生了一次更新。当所有条件都满足时，即边提灯人最终离开社区1时，定义这次更新发生。下文中，若 $\mathrm{P}(X = k) = (1-p)^k p$ for $k = 0,1,2,\ldots$，我们记作 $X \sim \mathrm{Geo}(p)$。那么，该更新过程的到达间隔时间分布为

$$\xi_n \overset{\text{dist.}}{=} \sum_{i=1}^{Z_1}(V_i + W_i + 2) + (Z_2 + 3),$$

其中

i. $\{V_i\}$ 是对应于第 $i-1$ 次和第 i 次访问社区1之间在社区1外所花时间的独立同分布的 Geo $\left(\frac{1}{2(K_n-1)}\right)$ 随机变量；

ii. $\{W_i\}$ 为独立同分布变量，并统计一次非更新访问期间在社区1内的移动次数（即边提灯人在第一次移动时不会移至社区1中的新顶点），由此可知，若边提灯人立即离开社区1（发生概率为 $1/2$），则为 $W_i = 0$；若边提灯人在紧接的下一步停留在当前节点（发生概率为 $(1/2)\cdot(1/n)$），则为 $W_i = 1 + W_i$；其中 $W_i \sim \mathrm{Geo}(1/2)$（注意，求和项 $+2$ 计入了进入和离开社区1的步骤）；

iv. $Z_1 \sim \mathrm{Geo}(1/2(1+1/n))$与 $\{V_i, W_i\}$相互独立，表示对社区1进行非更新访问的次数；也就是说，第一次移动是离开社区1或“停留”（即进入社区1后停留至紧接的下一步）；

v. $Z_2 \sim \mathrm{Geo}(1/2)$ 与 $\{V_i, W_i\}$ 和 Z_1 相互独立，表示边提灯人在更新前最后一次访问社区1时在社区1内部移动的次数（不包括首次移动到新顶点的那一步）（请注意，被加项 $+3$ 计算了进入社区1的步骤、在社区1中立即移动到新顶点的步骤，以及离开社区1的步骤）。

这里，

$$\mathbb{E}(\xi_n) = \frac{2n}{n+1}\left(2(K_n - 1) + \frac{3}{2n} + 2\right) + 5 \le 12K_n.$$

令 $R(t)$ 表示截至时间 t 的更新次数，并用 S_m 表示第 m 次更新时间，其中 $m = 0, 1, 2, \ldots\ldots$。令 $N = n^{1+a_2}\log^{3/2}(n)$ 和 $f(n) = n^{(3a_2+1)/2}\log^{3/2}(n)$，其中我们回顾一下，$0 < a_2 < 1$ 是定理选取的常数，用于调节社区1与其他社区之间的边概率。根据马尔可夫不等式，存在一个常数 $C > 0$，使得

$$\mathbb{P}\left(R(N) < f(n)\right) = \mathbb{P}\left(S_{f(n)} > N\right) \le \frac{\mathbb{E}(\xi_n) n^{(3a_2+1)/2}\log^{3/2}(n)}{n^{1+a_2}\log^{3/2}(n)} \le C n^{(a_2-1)/2}.$$

在两次更新之间，边提灯人必须至少进入社区1一次。这次进入来自一个均匀随机选取的社区，因为在社区层面上，随机游走是定义在 K 个社群上的简单对称随机游走。因此，一旦边提灯人离开社区1（这必须发生在两次更新之间），它将从另一个均匀随机选取的社区重新进入社区1。此外，进入社区与社区1之间随机选取的一条边和一条非边会发生切换。当从社区 i 移动到社区1时，社群之间边/非边的随机化等价于 Johnson图 $J(n_1n_i, m_{i1})$ [15]上的随机游走。这一随机游走的谱性质是众所周知的。特别地，此处定义的弛豫时间为 $1/(1-\lambda_2)$ [43],，其中 λ_2 是随机游走邻接矩阵的第二大特征值，且其等于 $t^{J,\,i}_{\text{rel}} = m_{i1}(n_in_1 - m_{i1})/n_in_1$。此外，由于 Johnson图上随机游走的平稳分布是均匀的，因此该随机游走的混合时间可按如下方式界定：

$$\binom{n_in_1}{m_{1i}} \le \frac{(en_1n_i)^{m_{1i}}}{m_{1i}^{m_{1i}}},$$

当 n 足够大时（其中 C 是一个可以随行改变的常数），

$$\begin{aligned} t^{J,i}_{\text{mix}}(\epsilon) &\le t^{J,i}_{\text{rel}} \log\left(\frac{(en_1n_i)^{m_{1i}}}{\epsilon m_{1i}^{m_{1i}}}\right) \le m_{1i}\log\left(\frac{(en_1n_i)^{m_{1i}}}{\epsilon m_{1i}^{m_{1i}}}\right) \\ &\le Cm_{1i}\log\left(\frac{(n_1n_i)^{m_{1i}}}{\epsilon}\right) \le Cm^*\log\left(\frac{(n_1n_i)^{m^*}}{\epsilon}\right). \end{aligned}$$

令 $t^{J*}_{\text{mix}}(\epsilon) = \max_i t^{J,\,i}_{\text{混合}}(\epsilon)$。设

$$\epsilon = (n_1n^*)^{-b_3'n^{a_2}},$$

我们有（再次令 C 为一个可以随行改变的常数）

$$t^{J*}_{\text{mix}}(\epsilon) = t^{J*}_{\text{mix}}\left((n_1n^*)^{-b_3'n^{a_2}}\right) \le Cn^{2a_2}\log(n_1n^*) \le Cn^{2a_2}\log(n).$$

一旦每个其他社区都已转移 $t^{J*}_{\text{mix}}(\epsilon)$ 次（即社区 2、3、.....、K_n 都已转移至社区 1），社区 1 与所有其他社区之间的信号便已均匀混合（混合程度为 ϵ）。随后，根据标准的优惠券收集者渐近性，对于适当的常数 C （该常数可以随行改变），经过

$$Cn^{(1-a_2)/2}n^{2a_2}\log(n)\,K_n\log K_n \le Cn^{(3a_2+1)/2}\log(n)$$

次更新后，我们以至少为 的概率访问了每个连接社区 1 的社区间网络（这里的社区间网络是由社区 1 与社区 $j \ne 1$ 之间的边 / 非边构成的 $K-1$ 个二部图），且访问每个网络至少 $t^{J*}_{mix}(\epsilon)$ 次

$$1 - Cn^{2a_2}\log(n)\,b_4^{-n^{(1-a_2)/2}}.$$

回忆一下，$\epsilon = (n_1n^*)^{b_3'n^{a_2}}$，并对每个整数 $\eta \in \mathbb{Z} > 0$，定义事件

$$\begin{aligned} \mathcal{A}_\epsilon(\eta) = \{&\text{For each } i \ne 1, \text{ there have been at least } t^{J*}_{\text{mix}}(\epsilon) \text{ transitions} \\ &\text{from community } i \text{ to community } 1 \text{ by time } \eta\} \end{aligned}$$

and

$\mathcal{B}(\eta) = \{$截至时间 $\eta\}$., 提灯人游走已经覆盖了社区1内的边集

在时刻 η, 如果满足以下条件，我们就称社区1内部以及社区1与所有其他社区之间的边配置是均匀随机的:

i 令 B_1 为社区1中的顶点集，则 $A_\eta[B_1] \sim$ ER(n_1, q) 与 $A_0[B_1]$ 相互独立；这里，$A_0[B_1]$ （分别为 $A_\eta[B_1]$ ）是 A_0 （分别为 A_η）在顶点 B_1 上的诱导子图;

ii. 对于 $i \neq 1$，令 B_i 为社区 i 中的顶点。令 $A_0[B_1;\ B_i]$ （分别为 $A_\eta[B_1;\ B_i]$）为 A_0 （分别为 A_η）在顶点集 B_1 和 B_i 之间的诱导二部子图，则 $A_\eta[B_1;\ B_i] \sim$ ER(n_1, n_i, m_{1i}) 与 $A_0[B_1$ 相互独立; $B_i]$; 这里，ER(n_1, n_i, m_{1i}) 是在 n_1 个顶点和 n_i 个顶点之间、含有 m_{1i} 条边的二部图上的均匀分布。此外，图 $A_\eta[B_1]$ $\{A_\eta[B_1;\ B_i]\}_{i\neq 1}^K$ 相互独立。

若情况如此，那么用 $\stackrel{\mathcal{D}}{=}$ 表示分布相等，我们有

$$\mathrm{Tr}\big(A_0(Q_1 \oplus I_{n-n_1})A_\eta(Q_1 \oplus I_{n-n_1})^T\big) \stackrel{\mathcal{D}}{=} \mathrm{Tr}\big(A_0(Q_2 \oplus I_{n-n_1})A_\eta(Q_2 \oplus I_{n-n_1})^T\big) \ \ \forall Q_1, Q_2 \in \Pi_{n_1}$$

令社区1内部及其与其他社区之间边的均匀分布记为 $\psi = \otimes_{i=1}^{K_n}\psi_i$—$\psi_1$ ，其中 ψ_1 是社区1内部的分布， ψ_i 是社区1与社区 i之间的分布。于是，与定理2的证明类似,

$$\mathbb{P}_\psi(\mathcal{E}_{\eta,\beta}) \leq (2n_1^\beta)^{1-n_1/(2n_1^\beta+1)}.$$

令 $\mathcal{C}$ 为社区1内部以及社区1与所有其他社区之间边配置的集合，并令

$$R^{(\eta)} = (R_1^{(\eta)}, R_2^{(\eta)}, \cdots, R_{K_n}^{(\eta)})$$

为取值于此类配置的随机变量，其中 $R_1^{(\eta)}$ 是社区1内部的配置， $R_i^{(\eta)}$ 是社区1与社区 i之间的配置，适用于 $i > 1$。于是我们有

$$\mathbb{P}(\mathcal{E}_{\eta,\beta}|\mathcal{A}_\epsilon(\eta), \mathcal{B}(\eta)) = \sum_{\vec{c}\ \in\ \mathcal{C}} \mathbb{P}\left[\mathcal{E}_{\eta,\beta}|R = c, \mathcal{A}_\epsilon(\eta), \mathcal{B}(\eta)\right] \mathbb{P}\left[R = c|\mathcal{A}_\epsilon(\eta), \mathcal{B}(\eta)\right].$$

现在，在给定 $\mathcal{A}_\epsilon$（η） $\mathcal{B}$（η）的条件下，我们有 η 大于以下总和

$$\eta \geq \tau_{cov,1} + \sum_{i=2}^{K_n} t_{mix}^{J,i}(\epsilon),$$

其中，$\tau_{cov,\ 1}$ 是社区1内所有无序对的覆盖时间。若我们以边提灯人在社区1内所走的步数（记为 η_1）以及在社区1与社区 i 之间所走的步数（记为 $\eta_{1,\ i}$）为条件，对所有满足以下条件的 $i \neq 1$

$$\eta_1 \geq \tau_{cov,1}; \quad \text{and} \quad \eta_{1,i} \geq t_{mix}^{J,i}(\epsilon) \text{ for each } i \neq 1; \tag{15}$$

那么 R 的坐标彼此独立。此外，记 P_R 为 R 的分布，我们有

$$\|P_R - \psi\|_{TV} = \|\otimes_{i=1}^{K-n} P_{R_i} - \otimes_{i=1}^{K-n}\psi_i\|_{TV} \leq \sum_{i=1}^{K_n}\|P_{R_i} - \psi_i\|_{TV} \leq K_n\epsilon.$$

由于此处全变差距离单调不增，且 $\|P_{R_1} - \psi_1\|_{TV}$是社区1内随机游走与平稳分布之间的距离；对于 $i \neq 1$， $\|P_{R_i} - \psi_i\|_{TV}$是社区1与社区 i之间随机游走（Johnson图）的距离与平稳分布之间的距离，因此我们有

$$\|P_R - \psi\|_{TV} = \leq \sum_{i=1}^{K_n}\|P_{R_i} - \psi_i\|_{TV} \leq K_n\epsilon.$$

由于此界对所有满足式15 的条件化（即社区1内部及跨社区的步数）均一致，因此我们有

$$|\mathbb{P}(R = c|\mathcal{A}_\epsilon(\eta), \mathcal{B}(\eta)) - \psi(c)| \leq K_n\epsilon,$$

从而

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_{\eta,\beta}|\mathcal{A}_\epsilon(\eta), \mathcal{B}(\eta)) &= \sum_{\vec{c}\in\mathcal{C}} \mathbb{P}(\mathcal{E}_{\eta,\beta}|R = c, \mathcal{A}_\epsilon(\eta), \mathcal{B}(\eta))\mathbb{P}(R = c|\mathcal{A}_\epsilon(\eta), \mathcal{B}(\eta)) \\
&\leq \sum_{\vec{c}\in\mathcal{C}} \mathbb{P}(\mathcal{E}_{\eta,\beta}|R = c)\psi(c) + K_n\epsilon \\
&\leq \mathbb{P}_\psi(\mathcal{E}_{\eta,\beta}) + K_n\epsilon \\
&\leq (2n_1^\beta)^{1-n_1/(2n_1^\beta+1)} + K_n\epsilon.
\end{aligned}$$

现在，令$\eta = N = n^{1+a_2}\log^{3/2}(n)$，且 $\epsilon = (n_1 n^*)^{-b_3' n^{a_2}}$ ，我们有

$$\begin{aligned}
\mathbb{P}(\mathcal{B}(N)) &\leq \mathbb{P}(\mathcal{B}(N)|R(N) \geq f(n)) + \mathbb{P}(R(N) < f(n)) \\
&\leq \mathbb{P}(\tau_{cov,1} > f(n)) + Cn^{a_2-1}/2 \\
&\leq e^{-\frac{n^{(3a_1+1)}\log^{3/2} n}{n^{2a_1}\log(n^{2a_1})}+1} + Cn^{a_2-1}/2 \\
&= e^{-\Omega(n^{a_1}\log^{1/2}(n))} + O(n^{a_2-1}/2)),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}(\mathcal{A}_\epsilon(N)) &\leq \mathbb{P}(\mathcal{A}_\epsilon(N)|R(N) \geq f(n)) + \mathbb{P}(R(N) < f(n)) \\
&O(n^{2a_2}\log(n)\, b_4^{-n^{(1-a_2)/2}}) + O(n^{a_2-1}/2)).
\end{aligned}$$

结合以上结果，我们有

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_{N,\beta}) &\leq \mathbb{P}(\mathcal{E}_{N,\beta}|\mathcal{A}_\epsilon(N), \mathcal{B}(N)) + \mathbb{P}(\mathcal{A}_\epsilon(N)) + \mathbb{P}(\mathcal{B}(N)) \\
&\leq (2n_1^\beta)^{1-n_1/(2n_1^\beta+1)} + K_n(n_1 n^*)^{-b_3' n^{a_2}} + O(n^{2a_2}\log(n)\, b_4^{-n^{(1-a_2)/2}}) \\
&\quad + O(n^{a_2-1}/2)) + e^{-\Omega(n^{a_1}\log^{1/2}(n))},
\end{aligned}$$

如所需。

7.6 证明块边提灯人的平稳分布

这里我们证明，平稳分布的形式为

$$\pi^\bullet(B_i,u,\vec{c}) \propto \frac{1}{K}\frac{1}{|B_i|}\prod_j q_{j;2}^{m_{j,\vec{c}}} q_{j;1}^{r_{j,\vec{c}}} \frac{1}{\prod_{i,j}\binom{n_i n_j}{m_{ij}}}$$

通过证明该　　分布满足细致平衡。我们考虑以下情形　　：

情形 1：　$u \neq v \in B_i$ 对于某个　B_i

i.　$\vec{c}_v = \vec{c}_u$ 此时　$\pi^\bullet(B_i,\ u,\ \vec{c}_u) = \pi^\bullet(B_i,\ v,\ \vec{c}_v)$，并且我们还有

$$\begin{aligned}\mathbb{P}_{(B_i,u,\vec{c}_u)}(W_1 = (B_i,v,\vec{c}_v) &= \mathbb{P}_{(B_i,v,\vec{c}_v)}(W_1 = (B_i,u,\vec{c}_u)\\ &= \begin{cases}\frac{1}{2}\frac{1}{|B_i|}(1-q_{i,1}) \text{ if } h_{G_t}(\{u,v\}) = 1\\ \frac{1}{2}\frac{1}{|B_i|}(1-q_{i,2}) \text{ if } h_{G_t}(\{u,v\}) = 0,\end{cases}\end{aligned}$$

因此满足细致平衡。

ii.　$\vec{c}_1 \neq \vec{c}_2$；不失一般性，设　$\vec{c}_u$　的　$\{u,v\}$　项为 0，　$\vec{c}_v$　的　$\{u,v\}$　项为 1（相反情形类似，留给读者处理）。在这种情况下，

$$\frac{\pi^\bullet(B_i,u,\vec{c}_u)}{\pi^\bullet(B_i,v,\vec{c}_v)} = \frac{q_{i;1}}{q_{i;2}}.$$

Now

$$\begin{aligned}\mathbb{P}_{(B_i,u,\vec{c}_u)}(W_1 = (B_i,v,\vec{c}_v)) &= \frac{1}{2}\frac{1}{|B_i|}q_{i,2};\\ \mathbb{P}_{(B_i,v,\vec{c}_v)}(W_1 = (B_i,u,\vec{c}_u)) &= \frac{1}{2}\frac{1}{|B_i|}q_{i,1}\end{aligned}$$

细致平衡关系也再次立即成立。

情形 2：对于某个　$B_j \neq B_i$，　$u \in B_i$、　$v \in B_j$。不失一般性，设　$\vec{c}_u$　和　$\vec{c}_v$　仅在　e，　e'（块　B_i　和　B_j 之间）处不同，其中 $h_{\vec{c}_u}(e) = 1$，　$h_{\vec{c}_u}(e) = 0$，且 $h_{\vec{c}_v}(e) = 0$，　$h_{\vec{c}_v}(e) = 1$。要从　$\vec{c}_u$ 转到　$\vec{c}_v$，选择移除　e　处的边，并添加　e'　处的非边，反之亦然。因此，

$$\begin{aligned}\mathbb{P}_{(B_i,u,\vec{c}_u)}(W_1 = (B_j,v,\vec{c}_v)) &= \frac{1}{2}\frac{1}{K-1}\frac{1}{|B_j|}\frac{1}{m_{ij}(n_i n_j - m_{ij})};\\ \mathbb{P}_{(B_j,v,\vec{c}_v)}(W_1 = (B_i,u,\vec{c}_u)) &= \frac{1}{2}\frac{1}{K-1}\frac{1}{|B_i|}\frac{1}{m_{ij}(n_i n_j - m_{ij})},\end{aligned}$$

并且满足细致平衡。

引理 7.7 的证明2

基于 V 中的 edgelighter 游走，考虑马尔可夫链 $W_t = (B_t,\ L_t,\ C_t)$，其中 L_t 表示时刻 t 边提灯人在 V 中的位置，B_t 表示 L_t 所属的社区，而 $C_t \in \{0,1\}^{\binom{n}{2}}$ 表示时刻 t 的边配置。令 $t^{\bullet}_{\mathrm{cov}}$ 为行走者遍历每个社区内所有边/非边的首次时刻（即，对于每个满足 $i,\ j$ 属于同一社区的 $\{i,j\} \in \binom{V}{2}$，行走者都已完成 $i \to j$ 转移和/或 $j \to i$ 转移）。令 η 为 W_0 的一个初始分布，使得

i 社区 B_0 是均匀随机选取的；

ii 给定 $B_0 = b$，L_0 在 b 上均匀分布；

iii 对于初始边配置 C_0，每个社区内部的所有边都不存在，而不同社区之间的所有边均独立且均匀随机（即，如果 $C^{i,\ j}$ 是块 i 和 j 之间的边的部分，那么 $C^{i,\ j}_0$ 相互独立，且 $\mathbf{P}(C^{i,\ j}_0 = c_0^{(i,j)}) = 1/\binom{n_i n_j}{m_{ij}}$）。

令 c' 为任意固定配置，使得在 c' 中每个社区内部的所有边都存在。给定 $\tilde{t}^{\bullet}_{\mathrm{cov}} \le t$，对于 $W_t = (b,\ x,\ c')$，每个社区内部的所有边都必须在时间 t 之前翻转（其概率为 $\prod_j q_{j;2}^{\binom{n_j}{2}}$）；并且给定社区 b，x 在 b 上均匀分布。这样的 W_t 不一定在每个社区上都赋予均匀质量，但存在一个社区 b 和一个顶点 $y \in b$，使得

$$\begin{aligned}\mathbb{P}_\eta(W_t = (b,x,c')) &= \mathbb{P}_\eta(W_t = (b,x,c') | \tilde{t}^{\bullet}_{\mathrm{cov}} \le t)\mathbb{P}(\tilde{t}^{\bullet}_{\mathrm{cov}} \le t) \\ &\le \mathbb{P}_\eta(\tilde{t}^{\bullet}_{\mathrm{cov}} \le t)\frac{1}{|K|}\frac{1}{|b|}\prod_j q_{j;2}^{\binom{n_j}{2}}\frac{1}{\prod_{i,j}\binom{n_i n_j}{m_{ij}}}.\end{aligned}$$

由此可得

$$1 - \frac{\mathbb{P}_\eta(W_t = (b,x,c'))}{\pi^{\bullet}(b,x,c')} \ge 1 - \mathbb{P}_\eta(\tilde{t}^{\bullet}_{\mathrm{cov}} \le t).$$

令 $s^{\bullet}(t,\eta) = \max_{(b,y,c)} 1 - \frac{\mathbb{P}^\eta(W_t=(b,y,c))}{\pi^{\bullet}(b,y,c)}$，并注意到 $\mathbf{P}(\tilde{t}^{\bullet}_{\mathrm{cov}} > t) \le s^{\bullet}(t,\eta)$。接下来，定义

$$\begin{aligned} d^{\bullet}(t,\eta) &= \|\mathbb{P}^t_\eta(\cdot) - \pi^{\bullet}\|_{TV} \\ d^{\bullet}(t) &= \max_{(b,x,c)} \|\mathbb{P}^t_{(b,x,c)}(\cdot) - \pi^{\bullet}\|_{TV} \\ \bar{d}^{\bullet}(t,\eta) &= \max_{(b,y,c)} \|\mathbb{P}^t_\eta(\cdot) - \mathbb{P}^t_{(b,y,c)}(\cdot)\|_{TV} \quad \text{and} \\ \bar{d}^{\bullet}(t) &= \max_{(b,x,c),(b',y,c')} \|\mathbb{P}^t_{(b,x,c)}(\cdot) - \mathbb{P}^t_{(b',y,c')}(\cdot)\|_{TV}. \end{aligned}$$

与引理 1 的证明中一样，我们有，由于 (Z_t) 是可逆的，

$$d^{\bullet}(t,\eta) \le \bar{d}^{\bullet}(t,\eta) \le d^{\bullet}(t,\eta) + d^{\bullet}(t) \quad \text{and} \quad s^{\bullet}(2t,\eta) \le 1 - (1 - \bar{d}^{\bullet}(t,\eta))^2,$$

其中，后一个不等式的控制 $s^{\bullet}(2t,\eta)$ 可按照 [3] 中引理 4.7 的证明得到。

虽然 $d^{\bullet}(t,\eta)$ 未必具有次乘性，但我们确实有 $d^{\bullet}(t+s,\eta) \le d^{\bullet}(s,\eta)d^{\bullet}(t)$。

这可由以下论证得出，该论证改编自 [43] 中引理 4.11 的证明：固定 (b, x, c)，并令 (U_s, V_s) 为 $\mathbb{P}^s_{(b,x,c)}(\cdot)$ 与 $\mathbb{P}^s_\eta(\cdot)$ 的最优耦合。于是，对于任意集合 A,

$$\begin{aligned}\mathbb{P}^{s+t}_{(b,x,c)}(A) - \mathbb{P}^{s+t}_\eta(A) &= \mathbb{E}(\mathbb{P}^t(U_s, A) - \mathbb{P}^t(V_s, A)) \\ &\le \mathbb{E}(\mathbb{1}\{U_s \neq V_s)\bar{d}^\bullet(t)) = \bar{d}^\bullet(t)\mathbb{P}(U_s \neq V_s) \le \bar{d}^\bullet(s,\eta)\bar{d}^\bullet(t),\end{aligned}$$

如所需。

令 $t^\bullet(\cdot)$ 为 W_t的通常混合时间。与引理 1的证明一样，注意到该马尔可夫链是可逆的，我们有：

$$\bar{d}^\bullet(2t^\bullet_m, \eta) \le \bar{d}^\bullet(t^\bullet_m, \eta)\bar{d}^\bullet(t^\bullet_m) \le 2(d^\bullet(t^\bullet_m, \eta) + d^\bullet(t^\bullet_m))d^\bullet(t^\bullet_m) \le 1/4.$$

因此，

$$\mathbb{P}_\eta(\tilde{t}^\bullet_{\text{cov}} > 4t^\bullet_m) \le s^\bullet(4t^\bullet_m, \eta) \le 1 - (1 - \bar{d}^\bullet(2t^\bullet_m, \eta))^2 < 1/2.$$

与引理1的证明一样，我们有

$$\mathbb{E}_\eta(\tilde{t}^\bullet_{\text{cov}}) \le 8t^\bullet_m.$$

无论起始位置在哪里，到时刻 $\tilde{t}^\bullet_{\text{cov}}$，$W_t$ 都必须已经遍历了最大社群内的所有 $\Theta(n^2)$ 条边/非边。因此，

$$\Omega(n^2) = \mathbb{E}_\eta(\tilde{t}^\bullet_{\text{cov}}) \le 8t^\bullet_m$$

如所需。

7.8 补充实验与图表

7.8.1 ER 模型上的补充标准边轻化过程图表

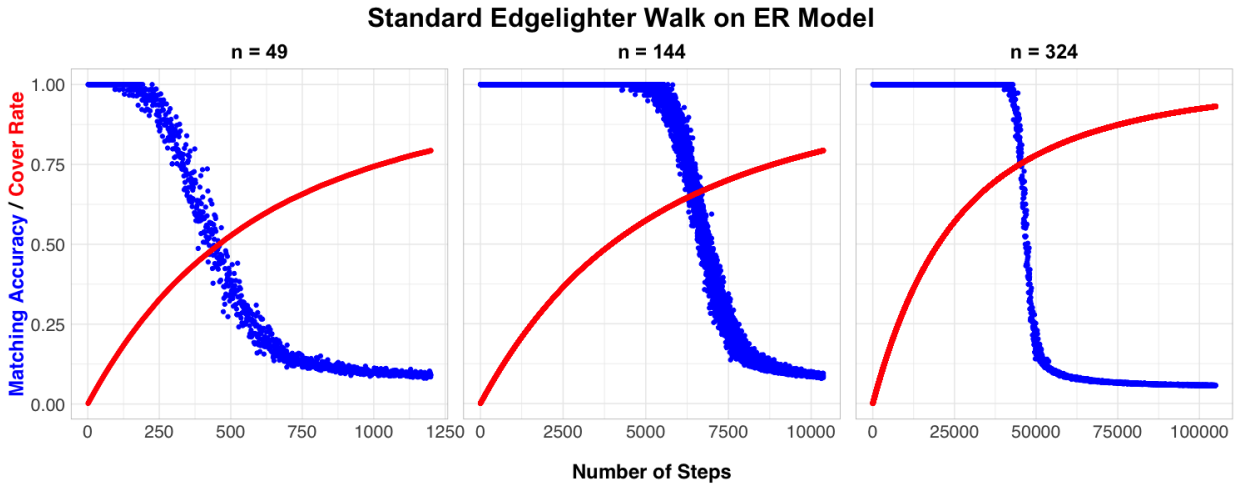

图 6：我们绘制了 $n = 49$ （左）、 $n = 144$（中）和 $n = 324$ （右）的匹配正确性随步数变化的曲线，颜色为蓝色；模型设置和所选参数与第 5.1 节相同。对于这些图表，我们还用红色表示边的覆盖率随步数的变化。

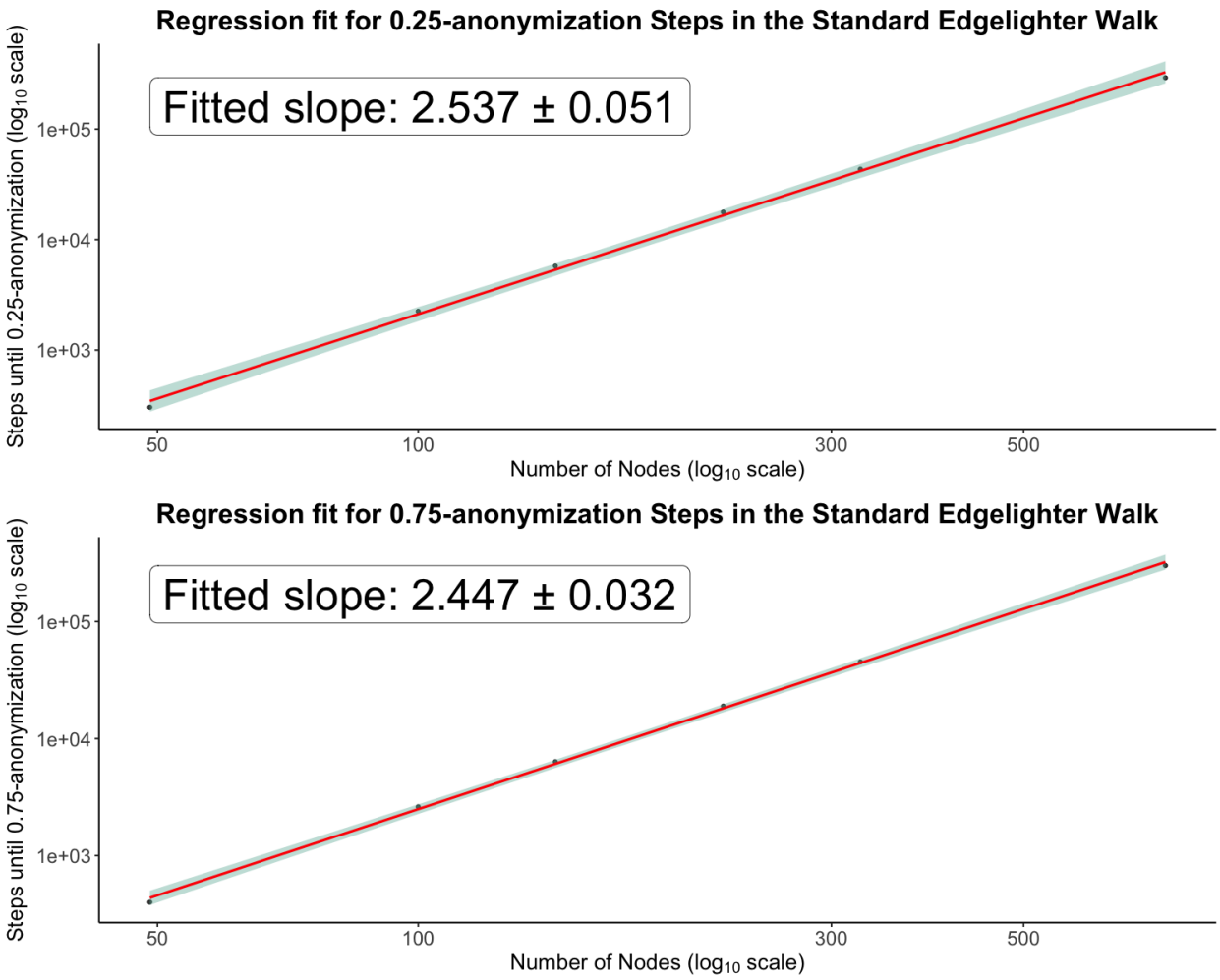

图 7：我们绘制了达到 0.25 匿名化所需步数（上图）以及达到 0.75 匿名化所需步数（下图）与节点数之间的回归关系 n，并采用双对数尺度。模型设置和所选参数与第5.1

7.8.2 随机块模型图中的附加区块边缘增强器

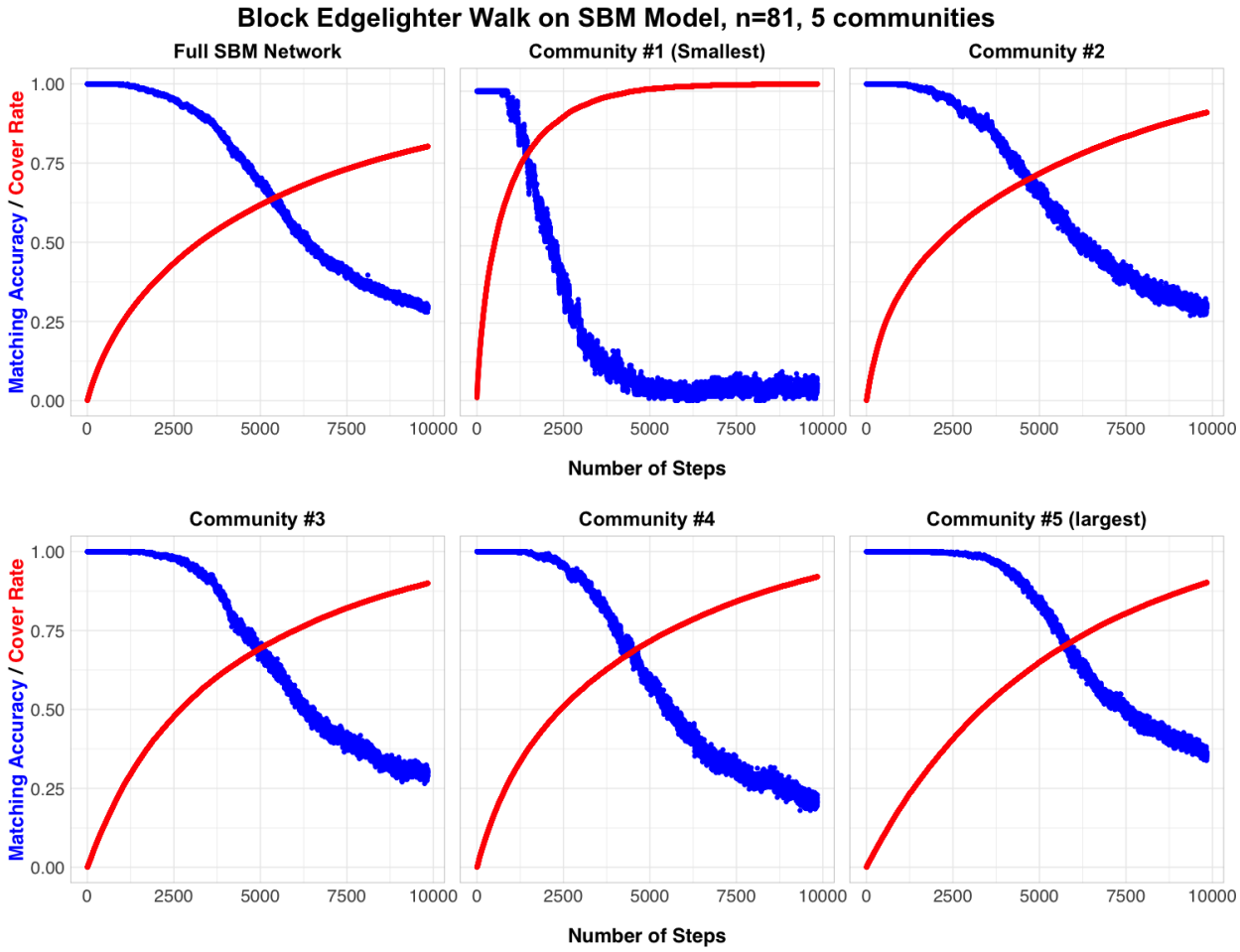

图8：随机块模型边提灯人游走图，包含 $n = 81$，实验设置的描述见第 5.1 节。社群数量 $= 5$

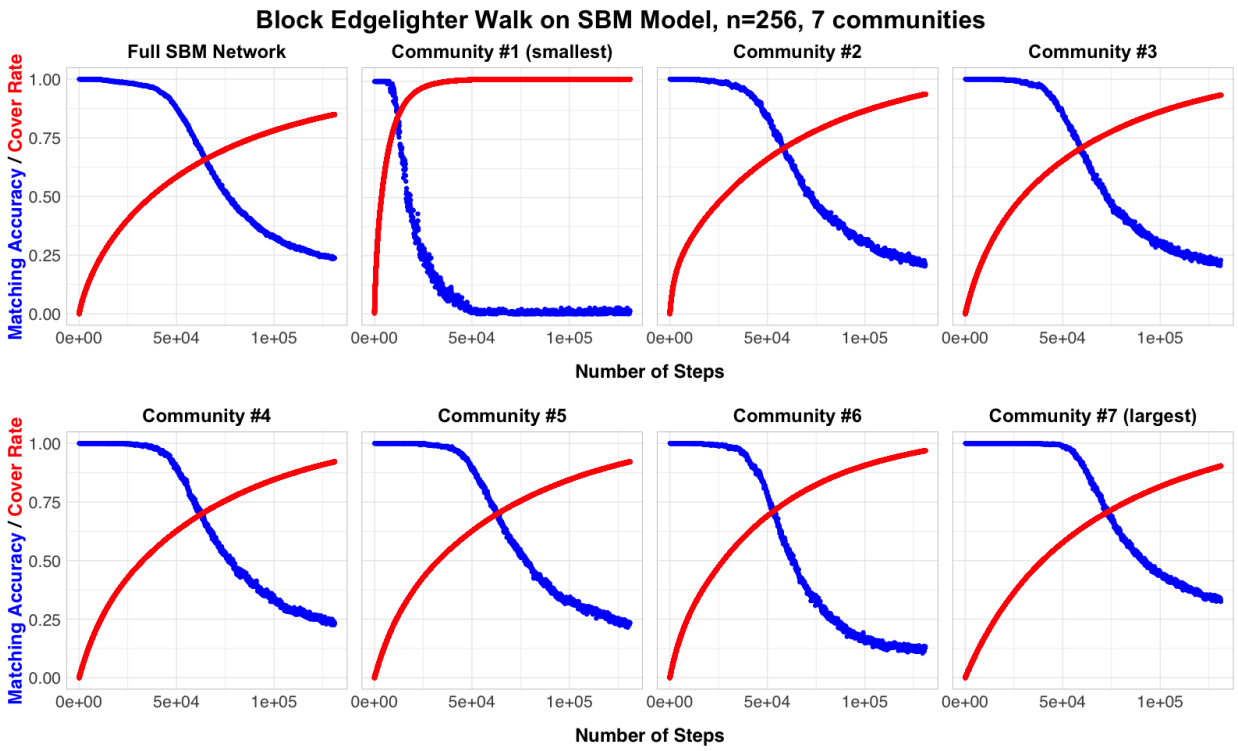

图 9：SBM edgelighter Walk 图，其中 $n = 256$，实验设置的说明请参见第 5.1 节。社群数量 $= 7$

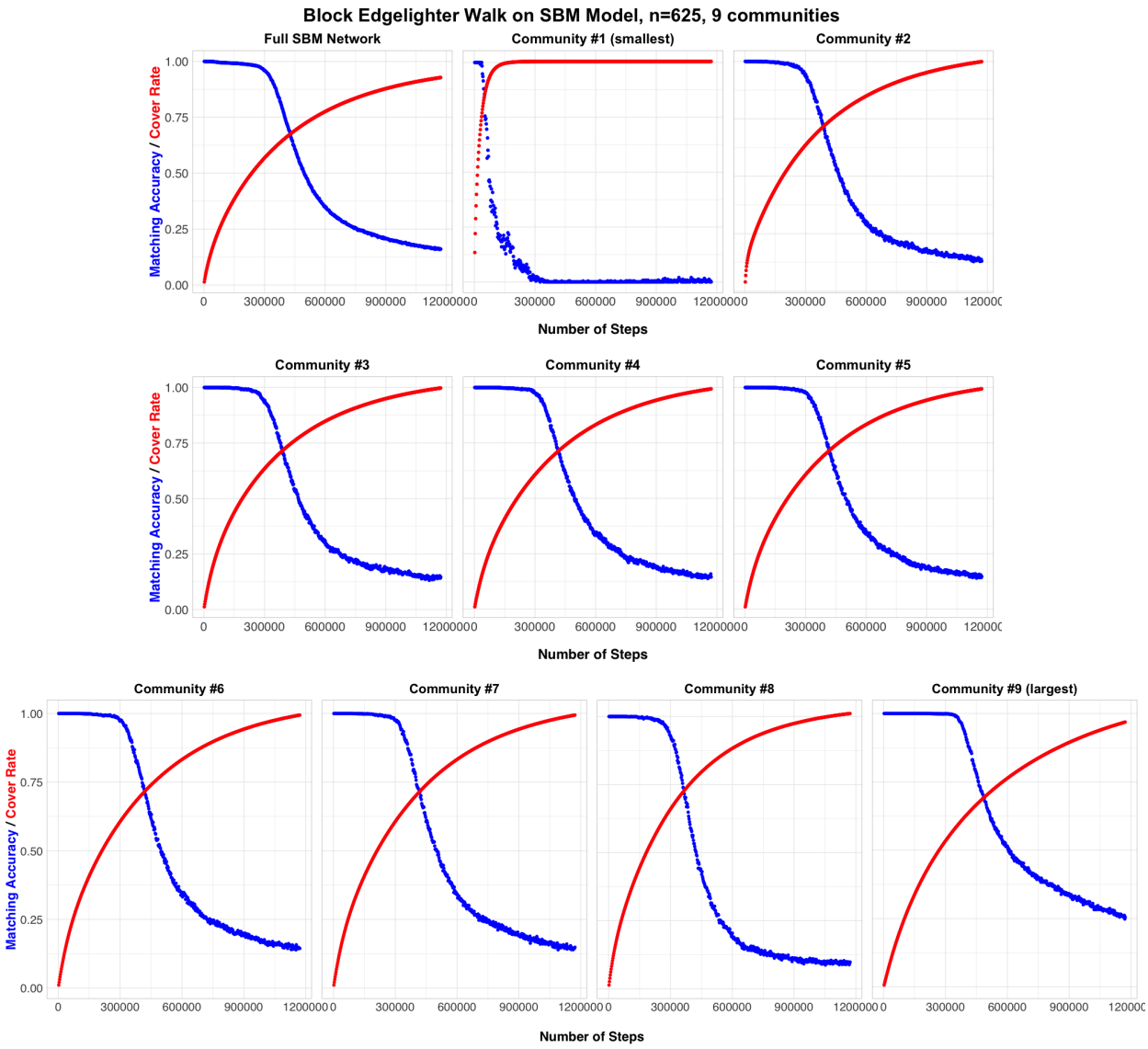

图10：SBM edgelighter Walk 图，其中包含 $n = 625$，请参见第 5.1 节，了解实验设置的描述。社群数量为 $= 9$

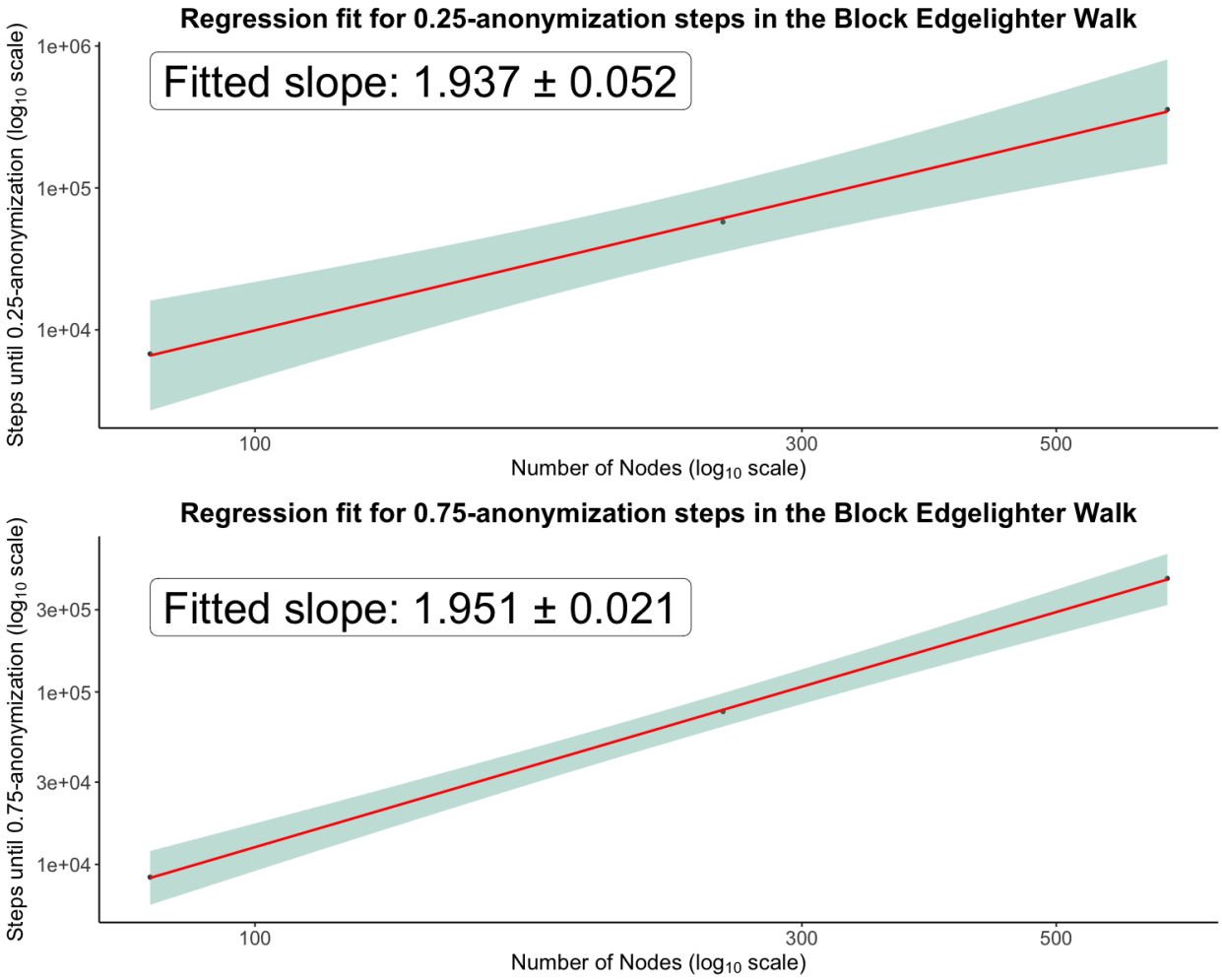

图11：我们绘制了达到 0.25-匿名化所需步数（上图）和达到 0.75-匿名化所需步数（下图）随节点数量变化的回归曲线 n ，采用双对数尺度。模型设置和所选参数与第5.1节相同。

.

7.8.3 EUemail 网络上的附加区块边缘增强器图 ts

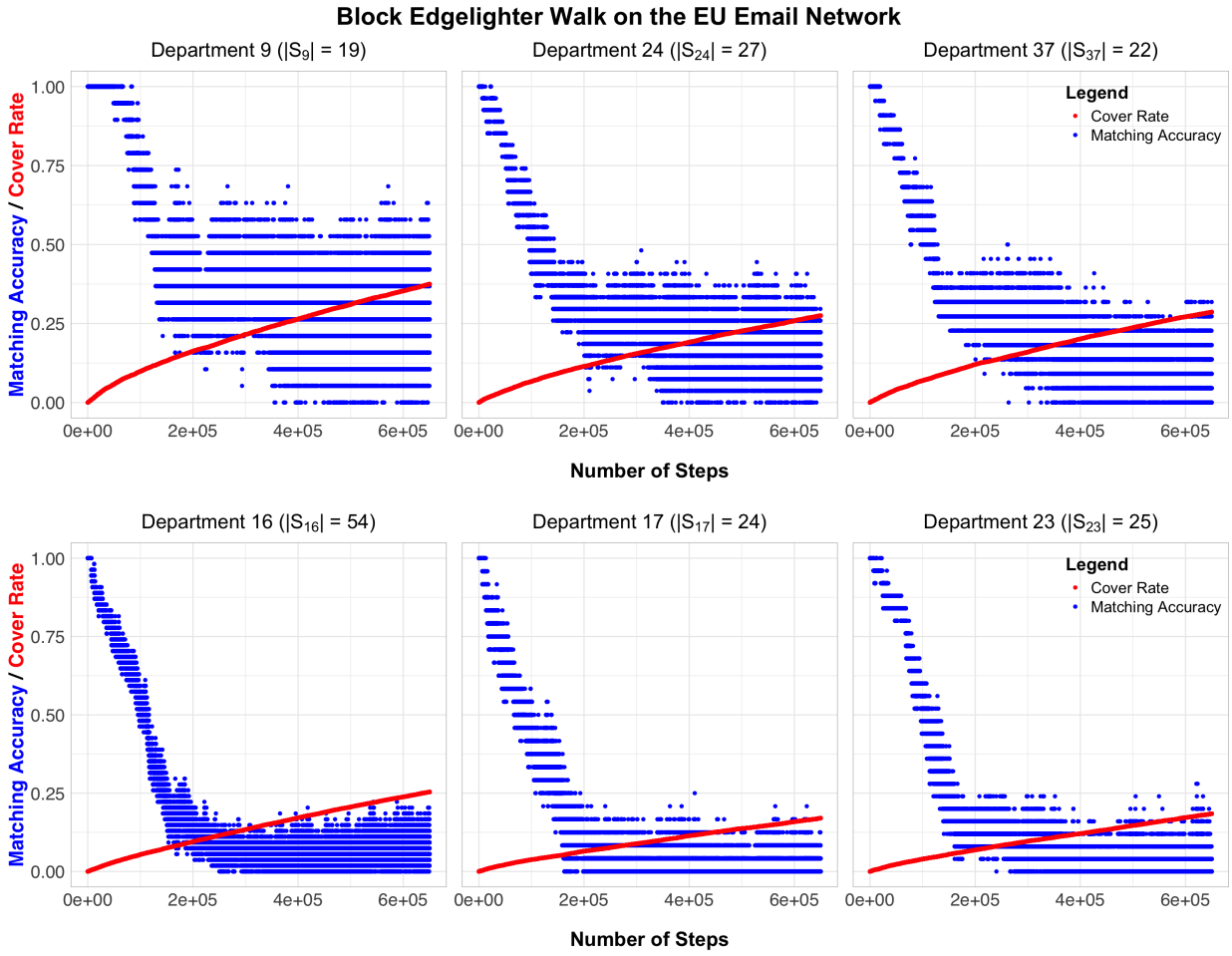

图12：EU email 网络数据的块边缘高亮游走图，实验设置的说明请参见第 5.2 节。