稀疏网络中易感-感染级联的最快推断

Anirudh Sridhar*、Tirza Routtenberg† 和 H. Vincent Poor*
*普林斯顿大学电气与计算机工程系，美国新泽西州普林斯顿，{anirudhs, poor}@princeton.edu
†内盖夫本-古里安大学电气与计算机工程系，以色列贝尔谢巴，tirzar@bgu.ac.il

arXiv:2302.05418v1 [math.ST] 10 Feb 2023

摘要——**我们研究如何尽可能快速地估计网络级联。假设该级联遵循一般的易感-感染过程，传播率因网络中的未知源点而异。尽管无法直接观测传播过程，但可以通过多轮易出错的诊断检测，获取有关其传播情况的噪声信息。我们提出一种新颖的自适应程序，能够在此观测模型下快速输出级联源点和完整传播范围的估计。值得注意的是，在网络拓扑满足温和条件时，即使在一个有n个顶点的网络中，级联影响到 多对数 n 个顶点之前，我们的程序也能估计级联的完整传播范围。我们还通过仿真结果补充理论分析，展示所提方法的有效性。**

I. 引言

大型网络往往容易受到级联故障的影响：异常行为从少数节点开始，并迅速扩散到网络的其余部分。如果不加以遏制，网络级联可能会对社会造成毁灭性影响。持续蔓延的 COVID-19疫情以及其他一些实例，都令人痛切地认识到这一点，包括虚假信息在社交网络中的传播 [7], [13], [31] ，以及恶意软件在信息物理网络中的传播 [16], [33]–[41]。因此，必须在过多节点遭到攻陷之前，准确定位网络级联。

遗憾的是，级联传播早期的信息往往带有噪声，这会使级联难以定位。举例来说，可以通过诊断检测来确定某一人群中的个体是否感染了传染病。正如早期用于检测 COVID-19 及其他疾病的抗原快速诊断检测准确性不高 [1], [9], 所表明的，这类检测可能有较高的假阳性率或假阴性率。在近期的一项研究中 [28]–[30], ，Sridhar 和 Poor 设计了用于确定网络级联源点的序贯估计器，在考虑这些不确定性的同时，还能迅速做出判断，以便及时采取缓解措施（例如隔离）。然而，他们的研究有一个严重缺陷：只有在极其简单的级联动力学和网络拓扑下，才证明了其估计量具备严格的性能保证。尤其是，在实际场景中，是否存在具有可证明性能保证的估计量，仍不明确。

在本研究中，我们通过设计一种新颖的序贯估计程序，来弥补这一空白，该程序适用于现实网络中有界度一般网络上的级联。我们假设级联由异质易感-感染（SI）过程建模，该过程能够描述多类型主体和病毒突变 [3], [4], [11], [32], [37]。随着级联传播，节点的行为会通过易出错的诊断检测定期监测。基于对这些含噪测量值的序贯观测，我们的程序能够快速输出准确的级联源点估计，以及级联的完整传播范围。值得注意的是，在对级联动力学和网络拓扑作出最少假设的情况下，我们证明，在一个 n 顶点图中，尚未有多对数数量的 n 个顶点受到影响时，就能可靠地估计级联的完整传播范围。我们通过模拟验证了这些理论结果，结果表明，在随机正则图中，当尚未有相当比例的节点受到影响时，我们的估计量就能快速定位 SI 级联。

A. 相关工作

我们的工作为网络级联的最快推断相关研究作出了贡献。这类研究的总体目标是利用网络拓扑和带噪声的顶点级信号，快速推断级联的相关特征。在 [42]–[44],中，Zou、Veeravalli、Li 和 Towsley 推导出了近乎最优的算法，用于检测级联在网络中缓慢传播时网络级联的出现。其初期工作的扩展也可参见 [23]–[25], [36], [39] 。与我们最相关的是 Sridhar 和 Poor[28]–[30],的工作，他们提出了相关方法，可在能够表示为格和正则树的网络中，快速估计简单确定性级联的源点。本文在 [30] 的方法基础上，设计了适用于有界度一般网络中现实级联的估计量。

我们还介绍了若干在不同背景下推断网络级联的有趣方向。在 Shah 和 Zaman 的开创性工作中， [26], [27], 研究者根据树中受影响顶点的大规模、无噪声快照来估计级联源（相比之下，本研究在一般网络中、在受影响顶点过多之前估计源点）。许多作者沿着[26], [27]最初的思路，在多个富有成果的方向上进行了拓展；例如参见 [12], [17], [34], [38], [40]。另一类研究通过布设少量无噪声“传感器”来检测和估计级联，从而获取关于级联影响的高质量测量值 [2], [6], [10], [14], [15], [18]。我们研究

一种互补情形：许多节点受到监测，但测量值可能相当嘈杂。

B. 论文结构

第 II 节介绍贯穿全文使用的符号约定。第 III 节详细介绍级联和诊断检测模型，以及用于估计级联源点和完整传播范围的自适应估计器。我们在第 IV 节从理论和实证两方面研究估计量的性能，并在第 V 节总结全文。主要结果的证明见附录。

II. 符号

我们分别用 $\mathbb{R}$、$\mathbb{R}_+$、$\mathbb{Z}$、$\mathbb{Z}_{\geq 0}$ 表示实数、正实数、整数和非负整数。对于有限集 S，$|S|$ 表示该集合中元素的数量。全文采用标准的渐近记号（例如，$O(\cdot)$ $o(\cdot)$）。此外，对于两个序列 $\{a_n\}_{n\geq 1}$ 和 $\{b_n\}_{n\geq 1}$，如果 $a_n \leq (1+o(1))\, b_n$，其中 $o(1) \to 0$ 当 $n \to \infty$ 时，我们称 $a_n \lesssim b_n$。

我们用一对 $G = (V(G)\ E(G))$ 表示一个图，其中 $V(G)$ 是该图的顶点集，$E(G)$ 是边集。图可以是有限的，也可以是无限的，这意味着顶点集的大小可以是有限的，也可以是无限的。顶点 v 的度记为 $\deg_G(v)$，即 v 在 G 中的邻居数量。图的最大度记为 Δ。对于任意 $u, v \in V(G)$，我们用 $u \sim v$ 表示 (u,v) 是 G 中的一条边。量 $\mathrm{dist}_G(u,v)$ 表示 u 和 v 在 G 中的最短路径距离。量 $\mathcal{N}_v(t)$ 是 v 在 G 中的 t 跳邻域，其中包含所有满足 $\mathrm{dist}_G(u,v) \leq t$ 的顶点 $u \in V(G)$。请注意，这一定义允许 t 取实数值，而不仅限于整数值。对于有限集 $S \subseteq V(G)$，我们定义 $\mathrm{diam}(S) := \max_{u,v\in S} \mathrm{dist}_G(u,v)$。

III. 模型与方法

A. 级联模型

我们使用网络中广为人知的 SI 过程的一般形式来模拟级联。在该模型中，级联以随机方式并在连续时间内传播。起初，级联仅由一个顶点 v^*（源点）组成。网络 G 中相邻的顶点会在随机时刻发生交互；当顶点 v 与其一个受影响邻居发生交互时，级联便会传播到顶点 v。

更具体地说，级联动力学可以描述如下。令 $\{\lambda_{uv}\}_{u\sim v}$ 为交互速率的集合，其中 $\lambda_{uv} \in \mathbb{R}_+$ 表示顶点 u 和 v 之间的交互频率。我们假设交互是对称的，因此 $\lambda_{uv} = \lambda_{vu}$。对于任意时间索引 $t \geq 0$，我们还将 $\mathcal{C}(t) \subset V$ 记为在时间 t 受到级联影响的顶点集合。当 $t = 0$ 时，$\mathcal{C}(0) = v^*$；换言之，初始时只有源点受到影响。对于任意 $t \geq 0$ 且 $\delta > 0$ 充分小时，对于任意顶点 $v \in V \setminus \mathcal{C}(t)$，有

$$\mathbb{P}(v \in \mathcal{C}(t+\delta) | \mathcal{C}(t)) = \delta \sum_{u\in\mathcal{C}(t):u\sim v} \lambda_{uv} + o(\delta), \qquad \text{(III.1)}$$

其中 $o(\delta) \to 0$ 的速率高于 $\delta \to 0$。换言之，未受影响顶点 $v \in V\setminus\mathcal{C}(t)$ 受到级联影响的速率，等于其与受影响邻居之间交互速率之和。

在所有交互速率都相等的特殊情况下（$\lambda_{uv} = \lambda$ 当 $u \sim v$ 时），SI 模型（III.1）在上个世纪受到了广泛关注。在数学物理学界，该模型被认为等价于边权服从 Exp(λ) 分布的首次通过渗流 [5], [22], ，也等价于无恢复过程的（马尔可夫）接触过程 [19], [20]。过程（III.1）还被用于推导广为人知的病毒传播群体层面模型（例如，参见 [8, 第 9] 章）。在（III.1）中采用异质交互速率，使我们能够刻画上述基本模型之外的多种重要情形。例如，一个人感染另一个人的速率可能显著取决于基础健康状况（例如，免疫功能受损者可能更容易感染）、佩戴口罩倾向以及病毒株类型。具体而言，动力学（III.1）与近期关于网络级联中的佩戴口罩和病毒突变的研究密切相关 [3], [4], [11], [32], [37]。[1]

级联的一个有用性质是，经过足够长的时间后，$\mathcal{C}(t)$ 可以包含在源点的两个邻域内。在陈述这一性质之前，我们定义 $\lambda_{min} := \min_{u\sim v} \lambda_{uv}$ 以及 $\lambda_{max} := \max_{u\sim v} \lambda_{uv}$。证明见附录 A。

命题 III.1。令 Δ 为 G 中的最大度。设

$$(\alpha, \beta) := \left(\frac{\lambda_{min}}{12 \log \Delta}, 3\Delta\lambda_{max} \right), \qquad \text{(III.2)}$$

并定义事件

$$\mathcal{E}_k := \{\forall t \geq k, \mathcal{N}_{v^*}(\alpha t) \subseteq \mathcal{C}(t) \subseteq \mathcal{N}_{v^*}(\beta t)\}.$$

则 $\lim_{k\to\infty} \mathbb{P}(\mathcal{E}_k) = 1$。

B. 观测模型

对于 t 的每个整数值以及每个 $v \in V$，我们假设以概率 p 对 v 进行易出错的诊断检测。诊断检测以概率 $1-\epsilon$ 给出正确结果（即，当 $v \in \mathcal{C}(t)$ 时输出值 1，并在 $v \notin \mathcal{C}(t)$ 时输出 -1）。以概率 ϵ，诊断检测结果不正确（包括误报和漏检错误）。形式化地，我们将 t 时刻对应于 v 的测量值记为 $Y_v(t)$，其中，如果不对 v 进行检测，则 $Y_v(t) = 0$；否则，$Y_v(t)$ 等于诊断检测的输出（-1 或 $+1$）。特别地，如果 $v \notin \mathcal{C}(t)$，则有

$$\begin{aligned} \mathbb{P}(Y_v(t) = 0 | \mathcal{C}(t)) &= 1-p, \\ \mathbb{P}(Y_v(t) = -1 | \mathcal{C}(t)) &= p(1-\epsilon), \\ \mathbb{P}(Y_v(t) = +1 | \mathcal{C}(t)) &= p\epsilon. \end{aligned} \qquad \text{(III.3)}$$

[1]这些研究考察了两个邻居之间的传播概率可能存在异质性的情况。由于个体之间的传播概率与传播率一一对应（例如参见 [21]），因此当恢复率为零时，此类模型与 (III.1) 密切相关。

作为简写，我们记作 $Y_v(t) \sim Q^-$，其中 Q^- 是 (III.3) 中描述的概率质量函数（PMF）。类似地，如果 $v \in \mathcal{C}(t)$，则有

$$\begin{aligned}\mathbb{P}(Y_v(t) = 0|\mathcal{C}(t)) &= 1-p,\\ \mathbb{P}(Y_v(t) = -1|\mathcal{C}(t)) &= p\epsilon,\\ \mathbb{P}(Y_v(t) = +1|\mathcal{C}(t)) &= p(1-\epsilon).\end{aligned} \tag{III.4}$$

在这种情况下，我们称 $Y_v(t) \sim Q^+$，其中 Q^+ 是式 (III.4) 中描述的概率质量函数。为简洁起见，我们还将 $\mathbf{Y}(t) := \{Y_w(t)\}_{w\in V(G)}$ 记作在正整数时刻 t收集到的所有信号的集合。

C. 源估计

任何估计程序都可以表示为一对 $(T, \hat{\mathbf{v}})$。其中，$T \in \mathbb{Z}_{\geq 0}$ 是一个取整数值、依赖于数据的停时，而 $\hat{\mathbf{v}} = \{\widehat{v}(t)\}_{t\geq 0}$ 是用于估计级联源的一系列估计量，其中 $\widehat{v}(t) \in V(G)$ 是数据的可测函数，即截至时刻 t的观测信号的可测函数。程序 $(T, \hat{\mathbf{v}})$ 会持续收集信号，直至达到停时T，此时输出 $\widehat{v}(T)$ 作为级联源的估计值。我们还假设存在一个有限的候选集 $U \subseteq V(G)$，表示已知的潜在源顶点集合（即已知 $v^* \in U$）。当 G 是有限图时，我们可以令 $U = V(G)$；不过，在理论结果中，我们也会考虑无限图。我们的目标是找到一个用于估计级联源的估计量 $\widehat{v}(T)$，使其估计误差较低（以到源点的图距离 $\mathsf{dist}_G(v^*, \widehat{v}(T))$ 衡量），同时确保尽可能快地作出决策（即 T 不宜过大），以防级联影响过多顶点。正如我们将在 III-D节中看到的那样，源估计还可以作为一个子程序，用于准确估计级联的完整传播范围。

我们的源点估计量基于两条指导原则：

1）靠近源点的顶点，其局部邻域中应有许多阳性病例。

2）如果 v 的邻域中的阳性病例数显著多于 u 的邻域中的阳性病例数，那么 v 更有可能在位置上靠近源点。

根据第一条指导原则，我们首先根据每个顶点的局部邻域中的阳性病例和阴性病例，为每个顶点构造一个 评分函数。对于顶点 $v \in U$ 和任意正整数时间索引 t，定义

$$Z_v(t) := \sum_{s=0}^{t} \sum_{w\in\mathcal{N}_v(\alpha s)} Y_w(s).$$

换言之，$Z_v(t)$ 是局部邻域 $\mathcal{N}_v(\alpha s)$ 中净阳性病例的累积和，其中 $0 \leq s \leq t$。参数 α 按照 (III.2) 选取，使得若 $v^* = v$，那么当s 足够大时，根据命题 III.1，$\mathcal{N}_v(\alpha s)$ 中的顶点很可能都受到级联的影响。另一方面，如果 $\mathsf{dist}_G(v^*, v) \geq (\alpha+\beta)t$ （其中 β 按照 (III.2) 定义），那么命题 III.1 意味着 $\mathcal{N}_v(\alpha t)$ 很可能与没有重叠。由此可知，当 $\mathsf{dist}_G(u, v) \geq (\alpha+\beta)t$ 时，通过 $Z_v(t) - Z_u(t)$ 衡量的评分函数之差，在估计源点时最具信息量。

根据第二条指导原则，我们的估计程序将持续收集数据，直到某个顶点 $v \in U$ 的评分显著大于所有与 v 距离足够远的顶点的评分。具体而言，对于每个 $v \in U$，我们定义停时 $T^{\alpha\beta}(v)$，它在满足以下条件的第一个时间 $t \geq 0$ 停止：

$$Z_v(t) - Z_u(t) \geq \frac{2\log|U|}{\log\left(\frac{1-\epsilon}{\epsilon}\right)}, \quad \forall u \in U : \mathsf{dist}_G(u, v) \geq (\alpha+\beta)t. \tag{III.5}$$

我们的估计程序正式表示为 $(T^{\alpha\beta}, \hat{\mathbf{v}})$，其中 $T^{\alpha\beta} := \min_{v\in U} T^{\alpha\beta}(v)$，且 $\widehat{v}(T^{\alpha\beta}) \in \arg\min_{v\in U} T^{\alpha\beta}(v)$。

注释 III.2。在对 $T^{\alpha\beta}(v)$ 的描述中，一个边界情况是：当 $t > \mathsf{dist}_G(u, v)/(\alpha+\beta)$ 时，(III.5) 中的条件是空泛的。因此，如果对于任意 $t \leq \mathsf{diam}(U)/(\alpha+\beta)$，所有停时均未停止，我们令 $T^{\alpha\beta} = \mathsf{diam}(U)/(\alpha+\beta)$，并令 $\widehat{v}(T^{\alpha\beta})$ 为 U 中的任意顶点。然而，正如我们的结果所表明的，这种边界情况发生的可能性非常低（见定理 *IV.2*）。

备注 III.3。在 (III.5) 中，$Z_v(t) - Z_u(t)$ 的阈值经过精心选择，使得 $\mathsf{dist}_G(v^*, \widehat{v}(T^{\alpha\beta})) \leq (\alpha+\beta)T^{\alpha\beta}$ 以趋近于 *1* 的概率成立，其中 $|U| \to \infty$。详情请参见引理 *C.2*。

注释 III.4。在 *[28], [30]* 中曾指出，源点估计可以视为一个序贯多假设检验问题；对于该问题，基于计算观测值的对数似然比的程序能够达到最优性能。事实上，在 $\mathcal{C}(t) = \mathcal{N}_{v^*}(t)$ 且 $\alpha = 1$, $Z_v(t) - Z_u(t)$ 的简单情形下，$Z_v(t) - Z_u(t)$ 与测度 $\mathbb{P}(\cdot|v^* = v)$ 和 $\mathbb{P}(\cdot|v^* = u)$之间的对数似然比成正比。然而，对于本文所考虑的更为现实的随机级联模型，计算对数似然比需要对级联的所有可能实现进行积分。我们转而使用 $Z_v(t) - Z_u(t)$ 作为对数似然比的替代量，从而规避了这些问题，同时还获得了与 *[28], [30]*中基于似然的方法相近的性能保证。

D. 从源估计到完整级联估计

另一个自然且重要的目标是快速估计级联的完整传播范围，而不仅仅是源点。利用过程（$T^{\alpha\beta}$, $\hat{\mathbf{v}}$）作为子程序，可以轻松实现这一目标。事实上，回顾命题 III.1 可知，如果 $T^{\alpha\beta}$ 足够大，那么 $\mathcal{C}(T^{\alpha\beta}) \subseteq \mathcal{N}_{v^*}(\beta T^{\alpha\beta})$ 以高概率成立。如果按照备注 III.3 的建议，$\mathsf{dist}_G(v^*, \widehat{v}(T^{\alpha\beta})) \leq (\alpha+\beta)T^{\alpha\beta}$，则可得

$$\mathcal{C}(T^{\alpha\beta}) \subseteq \mathcal{N}_{\widehat{v}(T^{\alpha\beta})}((\alpha+2\beta)T^{\alpha\beta}) =: \widehat{C}(T^{\alpha\beta}), \tag{III.6}$$

以高概率成立。因此，在这种情况下，级联传播范围的估计量 $\widehat{\mathcal{C}}(T^{\alpha\beta})$完全包含 真实传播范围。下文的定理 IV.2 验证了这一直觉。

IV. 性能分析

A. 理论结果

尽管我们的算法适用于任意 有界度图，但在理论结果中，我们聚焦于一类通用的无限 图。这类图刻画了这样一种情形：相对于整个网络的规模，级联较小，同时又不会使数学分析过于复杂，[2] 不过我们预计，对于直径足够大的有限图，我们的结果同样成立。形式上，我们作如下假设：

假设 IV.1. 假设 G 有无限多个顶点，且是连通的，并且最大度 Δ有限。

接下来定义几个关键量。对于非负整数 t 和顶点 $v \in V(G)$，定义 $f_v^\alpha(t) := \sum_{s=0}^{t} |\mathcal{N}_v(\alpha s)|$以及它的反函数 F_v^α。事实证明， $f_{v^*}^\alpha$和$F_{v^*}^\alpha$ 在我们估计量的性能保证中发挥着基础作用。事实上，根据评分函数的定义可以注意到，给定级联演化$\boldsymbol{\mathcal{C}}$ 的条件下：$= \{\mathcal{C}(s)\}_{s\geq 0}$,

$$\begin{aligned}\frac{\mathbb{E}[Z_{v^*}(t)|\boldsymbol{\mathcal{C}}]}{p(1-2\epsilon)} &= \sum_{s=0}^{t}(|\mathcal{N}_{v^*}(\alpha s)\cap\mathcal{C}(s)| - |\mathcal{N}_{v_*}(\alpha s)\setminus\mathcal{C}(s)|)\\ &= \sum_{s=0}^{t}|\mathcal{N}_{v^*}(\alpha s)| - 2\sum_{s=0}^{t}|\mathcal{N}_{v_*}(\alpha s)\setminus\mathcal{C}(s)| \sim f_{v^*}^\alpha(t),\end{aligned}$$

其中，依据命题 III.1，最后的渐近表达式以高概率成立。另一方面，如果 $u \in V(G)$满足 $\mathsf{dist}_G(v^*, u) > (\alpha+\beta)t$，那么在时刻 t， $\mathcal{N}_u(\alpha t)$中的顶点以高概率不受级联影响（见命题 III.1），因此 $\mathbb{E}[Z_u(t)|\boldsymbol{\mathcal{C}}] \leq 0$。由此可得， $\mathbb{E}[Z_{v^*}(t) - Z_u(t)|\boldsymbol{\mathcal{C}}] \geq p(1-2\epsilon)f_{v^*}^\alpha(t)$。如果$Z_v(t)$集中于其期望附近，那么当 $f_{v^*}^\alpha(t) \gtrsim \log|U|$时，我们看到它超过了（III.5）中的阈值；等价地，当 $t \lesssim F_{v^*}^\alpha(\log|U|)$时也是如此。下面的第一个主要结果证实了这一直觉。

定理 IV.2。 令(α, β) 依照(III.2) 设定，并令$U \subset V(G)$ 为任意有限的潜在源顶点候选集。那么，当 $|U| \to \infty$ 时，以下结论以趋近于 *1* 的概率成立：

1) $T^{\alpha\beta} \leq F_{v^*}^\alpha\left(\frac{15\log|U|}{p(1-2\epsilon)^2}\right)$

2) $\mathsf{dist}_G(v^*, \widehat{v}(T^{\alpha\beta})) \leq (\alpha+\beta)F_{v^*}^\alpha\left(\frac{15\log|U|}{p(1-2\epsilon)^2}\right)$

3) $\mathcal{C}(T^{\alpha\beta}) \subseteq \widehat{\mathcal{C}}(T^{\alpha\beta})$

第 1 项背后的直觉可参见定理陈述之前的讨论。特别要注意，随着测试频率降低（$p \to 0$）或测试误差增大（$\epsilon \to 1/2$）， $T^{\alpha\beta}$ 的上界会增大。第 2 项基本上可由阈值的选择推出，见（III.5）。如备注 III.3 所述，这些阈值确保 $\mathsf{dist}_G(v^*, \widehat{v}(T^{\alpha\beta}) \leq (\alpha+\beta)T^{\alpha\beta}$，而定理中的陈述可通过代入第 1 项给出的 $T^{\alpha\beta}$ 界限得出。第 3 项表明，级联估计量（定义见（III.6））不包含任何假阴性。完整的证明细节见附录 C。

在最一般的有界度图情形下，要对估计集合的大小给出界限颇具挑战。幸运的是，在图 G 的拓扑满足一个非常宽松的结构条件时，我们可以证明 $\widehat{\mathcal{C}}(T^{\alpha\beta})$ 的大小至多为 poly log $|U|$——相较于级联最初的$|U|$ 个潜在位置，这是指数级的缩减。以下推论对此进行了形式化；证明见附录 C。

推论 IV.3。 假设与定理 *IV.2* 相同的条件，并假设存在常数 q、 $r \geq 1$ 使得

$$|\mathcal{N}_u(t)| \leq q|\mathcal{N}_v(t)|^r, \qquad \forall u, v \in V(G), \forall t \geq 0. \tag{IV.1}$$

那么，当 $|U| \to \infty$ 时，概率趋于 *1*，存在常数 $c = c(\alpha, \beta, q, r)$，使得 $|\widehat{C}(T^{\alpha\beta})| \leq \log^c|U|$。

从高层次来看，条件（IV.1）表明邻域大小是多项式等价的，也就是说，一个邻域的大小可以由任意其他邻域的大小的某个（固定）多项式加以界定。该定义的一个推论是，如果某个邻域随 t 呈多项式增长（空间网络中的格等通常就是这种情况），那么所有 邻域都必须随 t 呈多项式增长。类似地，如果某个邻域呈指数增长（树状网络中通常就是这种情况），那么所有 邻域都必须呈指数增长。我们预计，在自然网络模型中，这一条件通常可以轻易满足。

最后，我们研究定理 IV.2 在简单网络中的含义，在这类网络中，邻域大小可以明确计算。值得注意的是，以下推论中关于 $T^{\alpha\beta}$ 的界与针对 [30] 中确定性级联推导出的最优源估计算法的性能相匹配。以下推论的证明可直接由定理 IV.2、引理 D.4 中对 F_v^α 的渐近刻画，以及附录 A] 中给出的格和正则树的邻域大小表达式得到。

推论 IV.4。 假设与定理 *IV.2* 中相同的条件，并进一步假设 G 是一棵度至少为 *3* 的无限正则树。以下陈述以趋于 *1* 的概率成立，当 $|U| \to \infty$ 时：

1) $T^{\alpha\beta} \lesssim \frac{1}{\alpha}\log\log|U|$

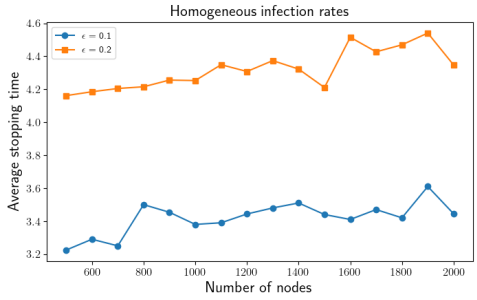
2) $\mathsf{dist}_G(v^*, \widehat{v}(T^{\alpha\beta})) \lesssim \frac{\alpha+\beta}{\alpha}\log\log|U|$

3) $\mathcal{C}(T^{\alpha\beta}) \subseteq \widehat{\mathcal{C}}(T^{\alpha\beta})$ *and* $|\widehat{\mathcal{C}}(T^{\alpha\beta})| \leq (\log|U|)^{3\beta/\alpha}$.

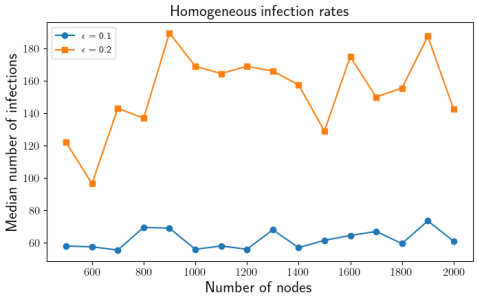
推论 IV.5. 假设满足定理 *IV.2* 中的相同条件，并进一步假设 G 是一个无限的 ℓ 维格。随着 $U \to \infty$ 趋于无穷，以下陈述以趋于 *1* 的概率成立：

1) $T^{\alpha\beta} = O\Big((\log|U|)^{1/(\ell+1)}\Big)$

[2]在有限图中，级联会在有限时间内传播至所有顶点，此时无法再从数据中获取有关源点位置的新信息。研究无限图则可以帮助我们避开此类“边界效应”。

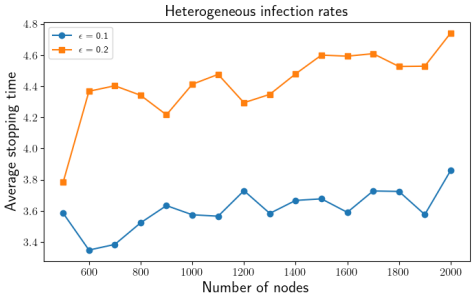


(a)

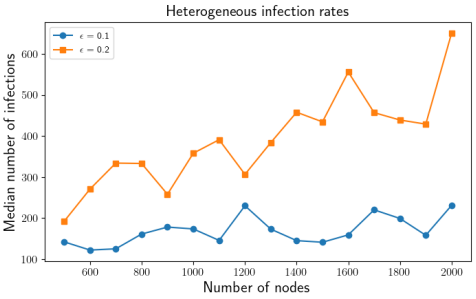


(b)

图1：均匀速率下期望停止时间（a）和停止时感染人数中位数（b）的曲线图。



(a)



(b)

图2：异质速率下期望停止时间（a）和停止时感染人数中位数（b）的曲线图。

2) $\mathrm{dist}_G(v^*, \widehat{v}(T^{\alpha\beta})) = O\Big((\log|U|)^{1/(\ell+1)}\Big)$

3) $\mathcal{C}(T^{\alpha\beta}) \subseteq \widehat{\mathcal{C}}(T^{\alpha\beta})$ *and* $|\widehat{\mathcal{C}}(T^{\alpha\beta})| = O\Big((\log|U|)^{\frac{\ell}{\ell+1}}\Big)$.

B. 仿真

为补充我们的理论结果，我们通过仿真表明，我们的级联估计器在非渐近情形（即 $|U|$ 取较小值时）也具有良好的性能。所用网络为一个均匀随机 3-正则图，规模介于 500 和 2000 个节点之间。每次试验中，级联源点均从网络中均匀随机选取。我们始终设定 $\alpha = 1$，$\beta = 2$，$p = 0.5$，并研究了 $\epsilon \in \{0.1, 0.2\}$。为生成图中的每个数据点，我们进行了 100 次独立试验，模拟级联传播和我们的估计程序。[3]

我们首先研究了估计量在一个（经典）SI模型中的表现，该模型具有均匀速率，其中 $\lambda_{uv} = 1$每当 $u \sim v$。我们的结果见图1，平均停时见图1a，感染次数的中位数见图1b。在所有仿真中，对于所有参数值，估计误差的平均值均小于1.5。我们绘制中位数而非平均值，以排除那些在网络中极快速传播的罕见级联实例。值得注意的是，当 $\epsilon = 0.1$ 时，即使网络规模不断增大，我们的估计量也能在受影响顶点达到80个之前可靠地追踪级联。尽管感染次数要多得多的是 $\epsilon = 0.2$，但这种情况下的平均停时似乎趋于平稳（见图1a），这表明当网络规模超过2000年后，感染曲线也应趋于平稳。

随后，我们研究了异质速率的情形，其中每个交互速率均取自 $[1, 1.5]$中的一个均匀随机变量。我们保持（α，β）=（1，2）。同样，对于所有测试的参数值，估计误差的平均值都相当小，不到2。有趣的是，速率异质性似乎并未显著影响估计量的性能，因为图1a和图2a中的平均停止时间相近。不过，由于级联传播速率的平均值有所增加，图2b中的感染人数确实有所上升。

[3]由于仿真中存在大量随机性（源自网络结构、级联演化和诊断检测模型），即使经过平均并取中位数，曲线仍可能显得嘈杂。

五、结论

在本研究中，我们考虑了如何根据网络级联传播过程的噪声时间序列来估计网络级联这一问题。针对这一情境下源估计的既有研究，只有在不切实际的简单级联和网络拓扑中才能给出可证明的性能保证 [28]–[30]。我们的工作具有广泛得多的适用性：我们为级联源点和完整传播范围开发了新型估计量，并为在任意有界度网络中传播的现实级联提供了可证明的保证。未来的研究方向包括：在我们考虑的一般情境下，研究用于估计级联源点和完整传播范围的最优 估计量；以及针对节点以自适应方式而非随机方式被选中接受诊断检测的情形，开发相应的估计量。

参考文献

[1]关于 COVID-19 即时免疫诊断检测的使用建议，2020年4月。网址：https://www.who.int/news-room/commentaries/detail/advice-on-the-use-of-point-of-care-immunodiagnostic-tests-for-covid-19.

[2] B. Adhikari、B. Lewis、A. Vulikanti、J. Jiménez 和 B. Prakash。医疗保健相关感染暴发的快速近最优监测。《PLOS计算生物学》，15(9)，2019年。[3] H. Alexander 和 T. Day。病原体演化出现的风险因素。英国皇家学会界面期刊，7 1474，2010 51 1455( ): – 。[4] A. Allard、P.-A. Noël、L. J. Dubé 和 B. Pourbohloul。多类型网络上的异质键渗流及其在流行病动力学中的应用。《物理评论E》，79:036113，2009年3月。[5] A. Auffinger、M. Damron 和 J. Hanson。《首次通过渗流50年》。美国数学学会，2017年。[6] P. Batlle、J. Bruna、C. Fernandez-Granda 和 V. M. Preciado。社会接触网络中用于疫情暴发检测与追踪的自适应检测分配。《工业与应用数学学会控制与优化杂志》，60(2):S274–S293，2022年。[7] A. Bessi、M. Coletto、G. A. Davidescu、A. Scala、G. Caldarelli 和 W. Quattrociocchi。科学与阴谋论：虚假信息时代的集体叙事。PLOS ONE，10(2):1–17，2015年2月。[8] F. Brauer 和 C. Castillo-Chavez。《种群生物学与流行病学数学模型》。施普林格，2012年。[9] A. Bruning、M. Leeflang、J. Vos、R. Spijker、M. de Jong、K. Wolthers 和 D. Pajkrt。流感、呼吸道合胞病毒及其他呼吸道病毒快速检测：系统综述与荟萃分析。65(6):1026–1032，2017年9月。[10] N. A. Christakis 和 J. H. Fowler。用于传染病暴发早期检测的社交网络传感器。PLOS ONE，5(9):1–8，2010年9月。[11] R. Eletreby、Y. Zhuang、K. M. Carley、O. Yağan 和 H. V. Poor。演化适应对复杂网络中传播过程的影响。《美国国家科学院院刊》，117(11):5664–5670，2020年。[12] G. Fanti、P. Kairouz、S. Oh 和 P. Viswanath。间谍对间谍：谣言源混淆。载于《2015年ACM SIGMETRICS计算机系统测量与建模国际会议论文集》，第271–284页，2015年。[13] A. Fourney、M. Z. Rácz、G. Ranade、M. Mobius 和 E. Horvitz。2016年美国总统选举期间假新闻消费的地理与时间趋势。载于《2017年ACM信息与知识管理会议论文集》，第2071–2074页，2017年。[14] A. Haldar、S. Wang、G. V. Demirci、J. Oakley 和 H. Ferhatosmanoglu。用于分析演化网络中传播的时序级联模型。ACM空间算法与系统汇刊，2023年1月。已接受，即将发表。[15] J. Heavey、J. Cui、C. Chen、B. Prakash 和 A. Vullikanti。网络流行病检测的可证明传感器集合与最小延迟。36(9):10202–10209，2022年。[16] J. O. Kephart 和 S. R. White。计算机病毒的有向图流行病模型。载于《1991年IEEE计算机学会安全与隐私研究研讨会论文集》，第343–359页，1991年5月。[17] J. Khim 和 P.-L. Loh。正则树中扩散源的置信集。《IEEE网络科学与工程汇刊》，4(1):27–40，2017年。[18] J. Leskovec、A. Krause、C. Guestrin、C. Faloutsos、J. VanBriesen 和 N. Glance。网络中高性价比的疫情暴发检测。载于《第13届ACM SIGKDD知识发现与数据挖掘国际会议论文集》，KDD '07，第420–429页，美国纽约州纽约，2007年。计算机协会。[19] T. Liggett。《随机相互作用系统：接触过程、投票过程与排斥过程》，第324卷。施普林格，柏林、海德堡，1999年。[20] D. Mollison。马尔可夫接触过程。《应用概率进展》，10(1):88–108，1978年。[21] M. E. J. Newman。网络中的流行病传播。《物理评论E》，66:016128，2002年。[22] D. Richardson。镶嵌中的随机生长。《剑桥哲学学会会刊》，74:515–528。[23] G. Rovatsos、G. V. Moustakides 和 V. V. Veeravalli。传感器网络中移动异常的最快检测。《IEEE信息论选定领域期刊》，2(2):762–773，2021年。

[24] G. Rovatsos、V. V. Veeravalli、D. Towsley 和 A. Swami。《网络中增长型动态异常的最快检测》。载于《2020年IEEE国际声学、语音与信号处理会议（ICASSP）论文集》，第8926–8930页，2020年。[25] G. Rovatsos、V. V. Veeravalli、D. Towsley 和 A. Swami。《传感器网络中位置和规模不断变化的异常的最快检测》。《IEEE航空航天与电子系统汇刊》，57(4)：2109–2120，2021年。[26] D. Shah 和 T. Zaman。《检测网络中的计算机病毒源点：理论与实验》。载于《2010年ACM SIGMETRICS计算机系统测量与建模国际会议论文集》，第203–214页，2010年。[27] D. Shah 和 T. Zaman。《网络中的谣言：谁是罪魁祸首？》。《IEEE信息论汇刊》，57(8)：5163–5181，2011年。[28] A. Sridhar 和 H. V. Poor。《网络级联的序贯估计》。载于《2020年第54届阿西洛马会议：信号、系统与计算机论文集》，第1507–1511页，2020年。[29] A. Sridhar 和 H. V. Poor。《寻找级联源点的贝叶斯最优方法》。载于《2021年IEEE国际声学、语音与信号处理会议（ICASSP）论文集》，第5190–5194页，2021年。[30] A. Sridhar 和 H. V. Poor。《利用噪声信息对网络级联进行最快推断》。即将发表于《IEEE信息论汇刊》，2023+。[31] E. Tacchini、G. Ballarin、M. L. Della Vedova、S. Moret 和 L. de Alfaro。《有人喜欢骗局：社交网络中的自动虚假新闻检测》。载于《第二届社会公益数据科学研讨会论文集》，第1–15页，2017年。[32] Y. Tian、A. Sridhar、O. Yağan 和 H. V. Poor。《佩戴口罩对病毒传播影响的分析：对COVID-19的启示》。载于《2021年美国控制会议（ACC）论文集》，第3132–3137页，2021年。[33] C. Wang、J. Knight 和 M. Elder。《计算机病毒感染及免疫的影响》。载于《第16届年度计算机安全应用会议（ACSAC'00）论文集》，第246–256页，2000年。[34] Z. Wang、W. Dong、W. Zhang 和 C. W. Tan。《利用多次观测进行谣言源头检测：基本极限与算法》。载于《2014年ACM SIGMETRICS计算机系统测量与建模国际会议论文集》，第1–13页，2014年。[35] D. Williams。《鞅与概率》。剑桥大学出版社，1991年。[36] X. Xian、A. Semenov、Y. Hu、A. Wang 和 Y. Jin。《大型网络中的自适应抽样与快速异常检测》。《IEEE自动化科学与工程汇刊》，第1–15页，2022年。[37] O. Yağan、A. Sridhar、R. Eletreby、S. Levin、J. Plotkin 和 H. Poor。《具有突变的多菌株模型下COVID-19传播的建模与分析》。《哈佛数据科学评论》，（特刊1），2021年。[38] L. Ying 和 K. Zhu。《大型网络中的扩散源定位》。《通信网络综合讲义》，11(1)：1–95，2018年。[39] R. Zhang、Y. Xie、R. Yao 和 F. Qiu。《使用扩散网络在线检测级联变化点》。载于《2022年第58届年度阿勒顿通信、控制与计算会议（阿勒顿会议）论文集》，第1–6页，2022年。[40] K. Zhu 和 L. Ying。《网络中的信息源检测：可能性与不可能性结果》。载于《第35届年度IEEE国际计算机通信会议论文集》，第1–9页，2016年。[41] C. C. Zou、W. Gong 和 D. Towsley。《红色代码蠕虫的传播建模与分析》。载于《第9届ACM计算机与通信安全会议论文集》，第138–147页，2002年。[42] S. Zou 和 V. V. Veeravalli。《传感器网络中动态事件的最快检测》。载于《2018年IEEE国际声学、语音与信号处理会议（ICASSP）论文集》，第6907–6911页，2018年。[43] S. Zou、V. V. Veeravalli、J. Li 和 D. Towsley。《结构化网络中重要事件的最快检测》。载于《2018年第52届阿西洛马会议：信号、系统与计算机论文集》，第1307–1311页，2018年。[44] S. Zou、V. V. Veeravalli、J. Li 和 D. Towsley。《网络中动态事件的最快检测》。《IEEE信息论汇刊》，66(4)：2280–2295，2020年。

附录 A 命题 III.1 的证明

在本附录中，我们通过事件 $\mathcal{E}_k$对级联的典型行为进行刻画。形式化地，对于正整数 k，定义事件

$$\mathcal{E}_{1,k} := \{\forall t \geq k, \mathcal{N}_{v^*}(\alpha t) \subseteq \mathcal{C}(t)\}$$
$$\mathcal{E}_{2,k} := \{\forall t \geq k, \mathcal{C}(t) \subseteq \mathcal{N}_{v^*}(\beta t)\},$$

其中，$(\alpha,\ \beta)$ 由 (III.2) 给出。本节的主要结果是以下两个引理。证明利用了级联的另一种表示形式（以指数边权表示），我们将在 A-A 节中对其进行介绍。

引理 A.1。对于任意正整数 k,

$$\mathbb{P}(\mathcal{E}_{1,k}^c) \leq \frac{12}{\lambda_{min}} e^{-k\lambda_{min}/12}.$$

引理 A.2。对于任意充分大的 k，有

$$\mathbb{P}(\mathcal{E}_{2,k}^c) \leq 900\left(\frac{2.9}{3}\right)^{3\Delta\lambda_{max}k}.$$

引理 A.1 和 A.2 的证明分别见 A-B 节和 A-C 节。此外，这两个引理可直接推出命题 III.1。

命题 *III.1* 的证明。注意到 $\mathcal{E}_k = \mathcal{E}_{1,\,k} \cap \mathcal{E}_{2,\,k}$。因此，对于任意 $v \in V(G)$，$\mathbb{P}(\mathcal{E}_k^c) \leq \mathbb{P}(\mathcal{E}_{1,\,k}^c) + \mathbb{P}(\mathcal{E}_{2,\,k}^c) \to 0$ 如$k \to \infty$，结论得证。□

A. 等价级联模型

我们介绍一种有用的替代表示法，用于描述 III-A 节中的级联动力学。先说明一些记号。生成一组独立随机变量$\{X_{uv}\}_{u\sim v}$，其中 $X_{uv} \sim \text{Exp}(\lambda_{uv})$ 表示边（$u,\ v$）的权重 。给定图中的一条路径[4] P，我们将该路径的权重（记为 weight(P)）定义为沿途各边权重之和。令 $\mathcal{P}_{vu}$ 表示所有从 v 出发并以 u为终点的路径集合。那么，在这个替代级联模型中，如果 v 是源点，则当且仅当 $\inf_{P\in\mathcal{P}_{vu}} \text{weight}(P) \leq t$ 时，$u \in \mathcal{C}(t)$。换言之，$\inf_{P\in\mathcal{P}_{vu}} \text{weight}(P)$ 表示 u的感染时间 ，其中v 是级联源点。由于指数分布具有无记忆性，这种表示法与（III.1）中的马尔可夫动力学等价。关于这一等价关系的详细说明，请参阅 [5, 第 6] 章及其中的参考文献。

B. 事件$\mathcal{E}_{1,k}$：引理 *A.1* 的证明

令 $u \in \mathcal{N}_{v^*}(\alpha s)$，并令 $d = \text{dist}_G(u,\ v)$。于是我们可以找到一条路径 $w_0, \ldots, w_d$，满足 $w_0 = v^*$和 $w_d = u$。注意到，u 的感染时间上界为 $\sum_{i=0}^{d-1} X_{w_i w_{i+1}}$，其中我们回顾一下，$X_{w_i w_{i+1}}$彼此独立，且 $X_{w_i,\ w_{i+1}} \sim \text{Exp}(\lambda_{w_i w_{i+1}})$。因此，对于任意 $s \geq 1$ 以及任意 $\theta \in (0,\ \lambda_{min})$，我们有

$$\begin{aligned}
\mathbb{P}(u \notin \mathcal{C}(s)) &\leq \mathbb{P}_v\left(\sum_{i=0}^{d-1} X_{w_i w_{i+1}} \geq s\right) \\
&\overset{(a)}{\leq} e^{-\theta s} \prod_{i=0}^{d-1} \frac{\lambda_{w_i w_{i+1}}}{\lambda_{w_i w_{i+1}} - \theta} \\
&= \exp\left\{-\theta s - \sum_{i=0}^{d-1} \log\left(1 - \frac{\theta}{\lambda_{w_i w_{i+1}}}\right)\right\} \\
&\overset{(b)}{\leq} \exp\left\{-\theta s - \alpha s \log\left(1 - \frac{\theta}{\lambda_{min}}\right)\right\}.
\end{aligned}$$

在 (a) 中，我们使用了指数随机变量的矩母函数形式；在 (b) 中，我们利用了函数 $x \mapsto \log(1-x)$ 对$x \in [0, 1)$递减，以及 $d \leq \alpha s$。令 $\theta = \lambda_{min}/2$，则对于$\alpha \leq \lambda_{min}/(6 \log 2)$，有

$$\mathbb{P}(u \notin \mathcal{C}(s)) \leq e^{-s\lambda_{min}/3}.$$

因此，我们还可以得到

$$\begin{aligned}
\mathbb{E}[|\mathcal{N}_{v^*}(\alpha s) \setminus \mathcal{C}(s)|] &= \sum_{u\in\mathcal{N}_{v^*}(\alpha s)} \mathbb{P}(u \notin \mathcal{C}(s)) \\
&\leq e^{-s\lambda_{min}/6} |\mathcal{N}_{v^*}(\alpha s)| \leq 2\left(e^{-\lambda_{min}/6}\Delta^{\alpha}\right)^s.
\end{aligned}$$

上式中，最后一个不等式使用了界 $|\mathcal{N}_{v^*}(\alpha s)| \leq 2\Delta^{\alpha s}$（见引理 D.1）。由马尔可夫不等式可得

$$\mathbb{P}(|\mathcal{N}_{v^*}(\alpha s) \setminus \mathcal{C}(s)| \geq 1) \leq 2\left(e^{-\lambda_{min}/6}\Delta^{\alpha}\right)^s \leq 2e^{-\frac{s\lambda_{min}}{12}},$$

其中，最后一个表达式是将 $\alpha = \lambda_{min}/(12 \log \Delta)$ 代入后得到的。最后，我们对所有 $s \geq k$ 应用并集界。□

C. 事件 $\mathcal{E}_{2,k}$：引理 *A.2* 的证明

我们首先证明一个关于独立指数随机变量之和的下尾界的有用中间结果。

引理 A.3. 设 $X_1, \ldots, X_m$ 是独立随机变量，且 $X_i \sim \text{Exp}(\mu_i)$，并令 $\mu := \max_{1\leq i\leq m} \mu_i$。那么对于任意 $\epsilon \in (0, \mu^{-1})$，当 m 足够大时，有

$$\mathbb{P}\left(\sum_{i=1}^{m} X_i \leq \epsilon m\right) \leq \sqrt{m}(2.8\epsilon\mu)^m.$$

证明。注意，如果 $\mu_1 \leq \mu_2$，其中 $X_1 \sim$ 服从 $\text{Exp}(\mu_1)$，且$X_2 \sim$ 服从 $\text{Exp}(\mu_2)$，那么 $X_2 \preceq X_1$。根据这一随机序，$\sum_{i=1}^m X_i$ 随机占优于一组独立同分布的 $\text{Exp}(\mu)$ 随机变量之和，该和的分布与$W \sim \text{Gamma}(m,\ \mu)$相同。因此

$$\mathbb{P}\left(\sum_{i=1}^{m} X_i \leq \epsilon m\right) \leq \mathbb{P}(W \leq \epsilon m) = \int_0^{\epsilon m} \frac{\mu^m x^{m-1} e^{-\mu x}}{(m-1)!} dx.$$

[4]路径是由互不相同的顶点组成的有限序列$u_1 \quad u_k \quad u_i \sim u_{i+1} \quad 1 \leq i \leq k-1$,...,, 满足对所有 。

注意， W 的密度在 $0 \le x \le (m-1)/\mu$ 上递增。因此，对于足够大的 $\epsilon < 1/\mu$ 和 m，我们可以将积分上界估计为

$$\frac{\mu^m(\epsilon m)^m e^{-\mu\epsilon m}}{(m-1)!} \sim \frac{\sqrt{m-1}}{e\sqrt{2\pi}}(\epsilon\mu e^{1-\mu\epsilon})^m \le \sqrt{m}(2.8\epsilon\mu)^m,$$

其中，该渐近展开源自斯特林公式。

□

现在我们可以证明主要结果了。

引理 *A.2* 的证明。令 $\mathcal{P}_d$ 为从 v^8 出发、长度为 d 的路径集合。注意，如果对所有 $P \in \bigcup_{d\ge\beta t}\mathcal{P}_d$，都有 $\mathsf{weight}(P) > t$，那么 $\mathcal{C}(t) \setminus \mathcal{N}_{v^*}(\beta t) = \emptyset$。因此，我们的证明将对路径 $P \in \bigcup_{d\ge\beta t}\mathcal{P}_d$ 的权重至多为 t 的概率进行界定。为此，我们有

$$\begin{aligned}
&\mathbb{P}(\mathcal{C}(t) \setminus \mathcal{N}_{v^*}(\beta t) \neq \emptyset) \\
&\le \mathbb{P}\left(\exists P \in \bigcup_{d\ge\beta t}\mathcal{P}_d : \mathsf{weight}(P) \le t\right) \\
&\le \sum_{d\ge\beta t} |\mathcal{P}_d| \max_{P\in\mathcal{P}_d} \mathbb{P}(\mathsf{weight}(P) \le t) \\
&\le \sum_{d\ge\beta t} |\mathcal{P}_d|\sqrt{d}\left(\frac{2.8\lambda_{max}}{\beta}\right)^d \\
&\le \sum_{d\ge\beta t} |\mathcal{P}_d|\sqrt{d}\left(\frac{2.8}{3\Delta}\right)^d, \qquad \text{(A.1)}
\end{aligned}$$

其中，第一个不等式由并集界得到；根据引理 A.3，当 t 充分大且满足 $\beta > \lambda_{max}$ 时，第三个不等式成立；最后一个表达式则由代入 $\beta = 3\Delta\lambda_{max}$ 得出。为简化最后的求和，我们将对 $\mathcal{P}_d$ 的大小进行界定。由于 G 中的最大度为 Δ，通过简单的计数可知，长度为 d 的路径至多有 Δ^d 条。当 t 充分大时，(A.1) 中的求和可界定为

$$\sum_{d\ge\beta t}\sqrt{d}\left(\frac{2.8}{3\Delta}\right)^d \le \sum_{d\ge\beta t}\left(\frac{2.9}{3\Delta}\right)^d = 30\left(\frac{2.9}{3}\right)^{\beta t}.$$

在上面的公式中，当 t 充分大时，不等式成立；最后一个表达式由几何级数求和公式得到。

综合以上结果，对于 k 充分大的情况，我们有

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_{2,k}^c) &\le \mathbb{P}(\exists t \ge k : \mathcal{C}(t) \setminus \mathcal{N}_{v^*}(\beta t) \neq \emptyset) \\
&\le \sum_{t\ge k}\mathbb{P}(\mathcal{C}(t) \setminus \mathcal{N}_{v^*}(\beta t) \neq \emptyset) \\
&\le \sum_{t\ge k} 30\left(\frac{2.9}{3}\right)^{\beta t} = 900\left(\frac{2.9}{3}\right)^{\beta k}.
\end{aligned}$$

□

附录 B 评分函数的性质

在本附录中，我们将以级联演化为条件，记作 $\boldsymbol{\mathcal{C}} := \{\mathcal{C}(s)\}_{s\ge0}$。我们还假设 $\boldsymbol{\mathcal{C}}$ 是满足事件 $\mathcal{E}_k$ 的级联实现（参见命题 III.1），其中 k取适当的值。我们还将定义测度 $\mathbb{P}_v(\cdot) = \mathbb{P}(\cdot|v^* = v)$。

我们首先建立评分函数之差的一个有用表示。令 $\{A_i\}_{i\ge1}, \{B_i\}_{i\ge1}$ 为相互独立的独立同分布 Q^+-分布随机变量集合。对于任意 $v \in V(G)$，令 $g_v^\alpha(t) := \sum_{s=0}^t |\mathcal{N}_v(\alpha s) \cap \mathcal{C}(s)|$。对于 $u, v \in V$ 以及任意 $t \ge 0$，定义过程

$$M_{vu}^\alpha(t) := \sum_{i=1}^{f_v^\alpha(t)+f_u^\alpha(t)} A_i \qquad \text{(B.1)}$$

$$D_{vu}^\alpha(t) := \sum_{i=1}^{f_v^\alpha(t)-g_v^\alpha(t)} (A_i + B_i). \qquad \text{(B.2)}$$

引理 B.1. 对于满足 $0 \le t \le \mathsf{dist}_G(v,u)/(\alpha+\beta)$ 的 t，

$$Z_v(t) - Z_u(t) \overset{d}{=} M_{vu}^\alpha(t) + D_{vu}^\alpha(t),$$

相对于测度 $\mathbb{P}_v(\cdot|\boldsymbol{\mathcal{C}})$。

证明。我们首先考虑$Z_v(t)$相对于 $\mathbb{P}_v(\cdot|\boldsymbol{\mathcal{C}})$ 的分布表示。根据信号模型的定义，对于任意 $t \ge 0$，有

$$\begin{aligned}
Z_v(t) &= \sum_{s=0}^t \left(\sum_{w\in\mathcal{N}_v(\alpha s)\cap\mathcal{C}(s)} Y_w(s) + \sum_{w\in\mathcal{N}_v(\alpha s)\setminus\mathcal{C}(s)} Y_w(s) \right) \\
&\overset{d}{=} \sum_{i=1}^{g_v^\alpha(t)} A_i - \sum_{i=1}^{f_v^\alpha(t)-g_v^\alpha(t)} B_i,
\end{aligned}$$

其中，$\{A_i\}_{i\ge1}$ 和 $\{B_i\}_{i\ge1}$ 是来自 Q^+的独立同分布样本集合。另一方面，注意到若事件 $\mathcal{E}_k$ 成立，则对于 $t \ge k$，$\mathcal{C}(t) \subseteq \mathcal{N}_v(\beta t)$，因此 $\mathcal{N}_u(\alpha t) \cap \mathcal{C}(t) = \emptyset$ 对于$u \in V_n \setminus \mathcal{N}_v((\alpha+\beta)t)$。由此可得，

$$Z_u(t) \overset{d}{=} -\sum_{i=1}^{f_u^\alpha(t)} A_i.$$

将这两个表达式相减，可得对于 $t \ge k$，

$$Z_v(t) - Z_u(t) \overset{d}{=} \sum_{i=1}^{g_v^\alpha(t)+f_u^\alpha(t)} A_i - \sum_{i=1}^{f_v^\alpha(t)-g_v^\alpha(t)} B_i.$$

令 $\{A_i'\}_{i\ge1}$ 为另一个由 Q^+分布的独立同分布随机变量序列，我们可以通过加减项得到

$$\begin{aligned}
Z_v(t) - Z_u(t) \overset{d}{=} &\left(\sum_{i=1}^{f_v^\alpha(t)-g_v^\alpha(t)} A_i + \sum_{i=1}^{g_v^\alpha(t)+f_u^\alpha(t)} A_i \right) \\
&- \sum_{i=1}^{f_v^\alpha(t)-g_v^\alpha(t)} (A_i' + B_i).
\end{aligned}$$

很容易看出，该表达式等于引理 B.1陈述中的表达式。 □

接下来，我们将证明 $M^{\alpha}_{vu}(t)$ 和 $D^{\alpha}_{vu}(t)$ 的一些有用性质。

引理 B.2. 在事件 $\mathcal{E}_k$ 上，几乎处处有 $\sup_{t\geq 0}|D^{\alpha}_{vu}(t)| \leq 2f^{\alpha}_{v}(k)$。

证明。 给定$\mathcal{C}$，根据$D^{\alpha}_{vu}(t)$ 的定义可得 $|D^{\alpha}_{vu}(t)| \leq 2(f^{\alpha}_{v}(t) - g^{\alpha}_{v}(t))$。在事件 $\mathcal{E}_k$ 上，对于所有 $t \geq k$，均有 $\mathcal{N}_v(\alpha t) = \mathcal{N}_v(\alpha t) \cap \mathcal{C}(t)$，因此

$$f^{\alpha}_{v}(t) - g^{\alpha}_{v}(t) = f^{\alpha}_{v}(k) - g^{\alpha}_{v}(k) \leq f^{\alpha}_{v}(k).$$

所需结论得证。 □

引理 B.3. 当 k 充分大时，对于 $t \geq 12k/(p(1-2\epsilon))$，有

$$\mathbb{P}_v\left(Z_v(t) - Z_u(t) \leq \frac{p(1-2\epsilon)}{3} f^{\alpha}_{v}(t)\middle|\mathcal{E}_k\right) \leq 2\exp\left(-\frac{p(1-2\epsilon)^2}{10} f^{\alpha}_{v}(t)\right).$$

证明。 给定$\mathcal{C}$，并假设事件 $\mathcal{E}_k$ 成立。简记 $N := f^{\alpha}_{v}(t) + f^{\alpha}_{u}(t)$。在 $\mathbb{P}_v$ 条件下，$M^{\alpha}_{vu}(t)$ 是由 N 个独立同分布的 Q^+ 分布随机变量之和。对于 $A \sim Q^+$，有 $\mathbb{E}[A] = p(1-2\epsilon)$，进而可推出$\mathbb{E}_v[M^{\alpha}_{vu}(t)|\mathcal{C}] = p(1-2\epsilon)N$。此外，$\mathrm{Var}(A) \leq \mathbb{E}[A^2] = p$ 也成立。因此，根据伯恩斯坦不等式可得

$$\mathbb{P}_v\left(M^{\alpha}_{vu}(t) \leq \frac{p(1-2\epsilon)}{2} N\middle|\mathcal{C}\right) \leq \exp\left(-\frac{p(1-2\epsilon)^2}{10} N\right). \tag{B.3}$$

现在我们利用 (B.3) 推导关于 $Z_v(t) - Z_u(t)$ 的不等式。我们可以写成

$$\begin{aligned}
&\mathbb{P}_v\left(Z_v(t) - Z_u(t) \leq \frac{p(1-2\epsilon)}{3} f^{\alpha}_{v}(t)\middle|\mathcal{C}\right) \\
&\overset{(a)}{\leq} \mathbb{P}_v\left(M^{\alpha}_{vu}(t) \leq \frac{p(1-2\epsilon)}{3} f^{\alpha}_{v}(t) + 2f^{\alpha}_{v}(k)\middle|\mathcal{C}\right) \\
&\overset{(b)}{\leq} \mathbb{P}_v\left(M^{\alpha}_{vu}(t) \leq \frac{p(1-2\epsilon)}{2} N\middle|\mathcal{C}\right) \\
&\overset{(c)}{\leq} \exp\left(-\frac{p(1-2\epsilon)^2}{10} f^{\alpha}_{v}(t)\right).
\end{aligned}$$

上式中，(a) 来自分解 $Z_v(t) - Z_u(t) = M^{\alpha}_{vu}(t) + D^{\alpha}_{vu}(t)$，并且根据引理 B.2，在事件$\mathcal{E}_k$ 上有 $|D^{\alpha}_{vu}(t)| \leq 2f^{\alpha}_{v}(k)$；$(b)$ 成立是因为 $N \geq f^{\alpha}_{v}(t)$ 假设 t 足够大，使得 $f^{\alpha}_{v}(k) \leq p(1-2\epsilon)f^{\alpha}_{v}(t)/6$，根据引理 D.2，当 $t \geq 12k/(p(1-2\epsilon))$ 时这一条件成立；而 (c) 则利用了 (B.3) 和 $N \geq f^{\alpha}_{v}(t)$。

为了将概率界与我们感兴趣的概率联系起来，对于任意事件 $\mathcal{A}$，我们可以写出：

$$\mathbb{P}_v(\mathcal{A}|\mathcal{E}_k) = \frac{\mathbb{P}_v(\mathcal{A}\cap\mathcal{E}_k)}{\mathbb{P}(\mathcal{E}_k)} = \frac{\mathbb{E}_v[\mathbb{P}_v(\mathcal{A}|\mathcal{C})\mathbf{1}(\mathcal{E}_k)]}{\mathbb{P}(\mathcal{E}_k)}.$$

令 $\mathcal{A} := \{Z_v(t) - Z_u(t) \leq p(1-2\epsilon)f^{\alpha}_{v}(t)/3\}$，并令 k 充分大，使得 $\mathbb{P}(\mathcal{E}_k) \geq 1/2$，即可得到所需结果。 □

引理 B.4. 固定 $u, v \in V(G)$，并令 T 为满足 $T \leq \mathrm{dist}_G(u,v)/(\alpha+\beta)$ 几乎处处成立的停时。那么，对于任意充分大的整数 k 和任意 $x \geq 0$，有

$$\mathbb{P}_v(Z_v(T) - Z_u(T) \leq -x|\mathcal{E}_k) \leq 2\left(\frac{\epsilon}{1-\epsilon}\right)^{x-2f^{\alpha}_{v}(k)}.$$

证明。 假设 $\mathcal{E}_k$ 成立。在给定 $\mathcal{C}$ 的条件下，我们声称过程 $(\epsilon/(1-\epsilon))^{M^{\alpha}_{vu}(t)}$ 关于 $\mathbb{P}_v(\cdot|\mathcal{C})$ 是一个鞅。为说明这一点，注意对于 $A \sim Q^+$，我们有

$$\mathbb{E}\left[\left(\frac{\epsilon}{1-\epsilon}\right)^{A}\right] = 1.$$

由于 $M^{\alpha}_{vu}(t)$ 是由 $f^{\alpha}_{v}(t) + f^{\alpha}_{u}(t)$ 个服从 Q^+ 分布的随机变量构成的独立同分布和，因此结论成立。

现在我们转而证明主要结果。设 T 是任意一个几乎处处有界的停时。那么，在给定 $\mathcal{C}$ 的条件下，我们有

$$\begin{aligned}
&\mathbb{P}_v(Z_v(T) - Z_u(T) \leq -x|\mathcal{C}) \\
&\overset{(a)}{\leq} \mathbb{P}_v(M^{\alpha}_{vu}(T) \leq -x + 2f^{\alpha}_{v}(k)|\mathcal{C}) \\
&= \mathbb{P}_v\left(\left(\frac{\epsilon}{1-\epsilon}\right)^{M^{\alpha}_{vu}(T)} \geq \left(\frac{1-\epsilon}{\epsilon}\right)^{x-2f^{\alpha}_{v}(k)}\middle|\mathcal{C}\right) \\
&\overset{(b)}{\leq} \left(\frac{\epsilon}{1-\epsilon}\right)^{x-2f^{\alpha}_{v}(k)} \mathbb{E}_v\left[\left(\frac{\epsilon}{1-\epsilon}\right)^{M^{\alpha}_{vu}(T)}\middle|\mathcal{C}\right] \\
&\overset{(c)}{=} \left(\frac{\epsilon}{1-\epsilon}\right)^{x-2f^{\alpha}_{v}(k)}.
\end{aligned}$$

上式中，(a) 源于引理 B.1 中 $Z_v(T) - Z_u(T)$ 的表示（该表示成立是因为 $T \leq \mathrm{dist}_G(u,v)/(\alpha+\beta)$ 几乎处处），以及引理 B.2 中得出的 $D^{\alpha}_{vu}(T)$ 界；(b) 由马尔可夫不等式得出；而 (c) 则是可选停时定理的推论 [35, 第 10 节。10]，因为 T 几乎处处有界。

最后，为了将以级联为条件改为以 $\mathcal{E}_k$ 为条件，我们可以采用与引理 B.3 的证明相同的推理。 □

附录 C 定理 IV.2 和推论 IV.3 的证明

我们将证明一系列引理，以确立 $T^{\alpha\beta}$、$\mathrm{dist}_G(v^*, \widehat{v}(T^{\alpha\beta}))$ 和 $\widehat{\mathcal{C}}(T^{\alpha\beta})$ 的性质。定理 IV.2 的证明可由这些结果直接推出，详见本节末尾。

引理 C.1. 设 k 足够大，且 $v \in U$。那么，当 $|U|$ 足够大时，

$$\mathbb{P}\left(T^{\alpha\beta} \geq F^{\alpha}_{v^*}\left(\frac{15\log|U|}{p(1-2\epsilon)^2}\right)\middle|\mathcal{E}_k\right) \leq 2|U|^{-1/2}.$$

证明。 定义量

$$t_v := F_v^{\alpha}\left(\frac{15\log|U|}{p(1-2\epsilon)^2}\right), \tag{C.1}$$

并注意到

$$\frac{p(1-2\epsilon)}{3}f_v^{\alpha}(t_v) \geq \frac{5\log|U|}{1-2\epsilon} \geq \frac{5\log|U|}{\log\left(\frac{1-\epsilon}{\epsilon}\right)} \geq \tau. \tag{C.2}$$

此外，我们可以将所关注的概率界定为

$$\begin{aligned}
&\mathbb{P}_v(T^{\alpha\beta} \geq t_v|\mathcal{E}_k) \leq \mathbb{P}_v(T^{\alpha\beta}(v) \geq t_v|\mathcal{E}_k)\\
&\overset{(a)}{\leq} \sum_{u\in U\setminus\mathcal{N}_v((\alpha+\beta)t_v)} \mathbb{P}_v(Z_v(t_v) - Z_u(t_v) < \tau|\mathcal{E}_k)\\
&\overset{(b)}{\leq} 2\exp\left(\log|U| - \frac{p(1-2\epsilon)^2}{10}f_v^{\alpha}(t_v)\right)\\
&\overset{(c)}{\leq} 2\exp\left(-\frac{1}{2}\log|U|\right).
\end{aligned}$$

上式中，(*a*)源于并集界，(*b*)是引理 B.3 的推论，因为 (C.2)成立，而(*c*)使用了t_v的定义。 □

引理 C.2。 令 k 充分大。则对于任意 $v \in U$，有

$$\mathbb{P}_v\big(\mathsf{dist}_G(v,\widehat{v}^{\alpha\beta}) > (\alpha+\beta)T^{\alpha\beta}\big|\mathcal{E}_k\big) \leq c_{k,\epsilon}|U|^{-1},$$

其中$c_{k,\,\epsilon}$ 是一个仅依赖于 k 和 ϵ的常数。

证明。 我们首先写出

$$\begin{aligned}
&\mathbb{P}_v\big(\mathsf{dist}_G(v,\widehat{v}^{\alpha\beta}) \geq (\alpha+\beta)T^{\alpha\beta}\big|\mathcal{E}_k\big)\\
&\quad= \sum_{u\in U}\mathbb{P}_v\left(\widehat{v}^{\alpha\beta}=u, T^{\alpha\beta} < \frac{\mathsf{dist}_G(v,u)}{\alpha+\beta}\middle|\mathcal{E}_k\right).
\end{aligned} \tag{C.3}$$

为分析求和中的概率，我们定义停时 $T_{vu} := \mathsf{max}\{T^{\alpha\beta}, \mathsf{dist}_G(v,u)/(\alpha+\beta)\}$。于是我们可以得到上界

$$\begin{aligned}
&\mathbb{P}_v\left(\widehat{v}^{\alpha\beta}=u, T^{\alpha\beta} < \frac{\mathsf{dist}_G(v,u)}{\alpha+\beta}\middle|\mathcal{E}_k\right)\\
&\leq \mathbb{P}_v(Z_v(T_{vu}) - Z_u(T_{vu}) \leq -\tau|\mathcal{E}_k)\\
&\leq \left(\frac{\epsilon}{1-\epsilon}\right)^{\tau-f_v^{\alpha}(k)} =: \left(\frac{1-\epsilon}{\epsilon}\right)^{f_v^{\alpha}(k)}|U|^{-2}.
\end{aligned} \tag{C.4}$$

上文中，第一个不等式成立是因为 $T_{vu} = T^{\alpha\beta}$ when $T^{\alpha\beta} < \mathsf{dist}_G(u,v)/(\alpha+\beta)$，第二个不等式则由引理 B.4 得出。在最后一个表达式中，我们代入 $\tau = 2\log|U|/\log((1-\epsilon)/\epsilon)$。最后，将 (C.4) 中的上界代入 (C.3)，可知我们关注的概率至多为 $c_{k,\epsilon}|U|^{-1}$，其中 $c_{k,\epsilon} := ((1-\epsilon)/\epsilon)^{f_v^{\alpha}(k)}$。 □

引理 C.3。回顾

$$\widehat{\mathcal{C}}(T^{\alpha\beta}) := \mathcal{N}_{\widehat{v}(T^{\alpha\beta})}((\alpha+2\beta)T^{\alpha\beta}).$$

对于足够大的 k，以及所有远大于 k 的 $|U|$，有

$$\mathbb{P}\Big(\mathcal{C}(T^{\alpha\beta}) \not\subseteq \widehat{\mathcal{C}}(T^{\alpha\beta})\Big|\mathcal{E}_k\Big) \leq c_{k,\epsilon}|U|^{-1},$$

其中 $c_{k,\epsilon}$ 是引理 *C.2* 中的常数。

证明。 注意，对于固定的 k，当 $T^{\alpha\beta} \geq k$ 足够大时，有 $|U|$。事实上，对于任意 u，$v \in U$，有$|Z_v(t) - Z_u(t)| \leq f_v^{\alpha}(t) + f_u^{\alpha}(t)$，因为所有观测信号的绝对值都不超过 1。因此，对于所有 $0 \leq t \leq k$，$|Z_v(t) - Z_u(t)| \leq f_v^{\alpha}(k) + f_u^{\alpha}(k)$，当 $|U|$ 足够大时，这小于阈值 $2\log|U|/\log((1-\epsilon)/\epsilon)$。由此立即可知，$T^{\alpha\beta}$ 必须大于 k。

接下来，当 k 足够大时，我们有

$$\begin{aligned}
&\mathbb{P}_v\Big(|\mathcal{C}(T^{\alpha\beta}) \setminus \widehat{\mathcal{C}}(T^{\alpha\beta})| \geq 1\Big|\mathcal{E}_k\Big)\\
&\overset{(a)}{\leq} \mathbb{P}_v\big(\exists u \in \mathcal{N}_v(\beta T^{\alpha\beta}) \setminus \mathcal{N}_{\widehat{v}^{\alpha\beta}}((\alpha+2\beta)T^{\alpha\beta})\big|\mathcal{E}_k\big)\\
&\overset{(b)}{\leq} \mathbb{P}_v\big(\mathsf{dist}_G(v,\widehat{v}^{\alpha\beta}) \geq (\alpha+\beta)T^{\alpha\beta}\big|\mathcal{E}_k\big)\\
&\overset{(c)}{\leq} c_{k,\epsilon}|U|^{-1}.
\end{aligned}$$

上式中，(*a*) 成立是因为当 v 为源点时，对于 $\mathcal{E}_k$ 上所有的 $t \geq k$，$\mathcal{C}(t) \subseteq \mathcal{N}_v(\beta t)$；(*b*) 是由三角不等式得到的；而 (*c*) 则由引理 C.2 得出。 □

定理 *IV.2* 的证明。注意，对于固定但足够大的 k，引理 C.1、C.2 和 C.3 中的概率界在 $|U| \to 0$ 时趋于零。再注意到，根据命题 III.1，$\mathbb{P}(\mathcal{E}_k) \to 1$，因此我们可以令 k 随 $|U|$ 以任意缓慢的速度增长，从而消去引理 C.1、C.2 和 C.3 中对 $\mathcal{E}_k$ 的条件限制。由并集界即可得到该定理。

□

推论 *IV.3* 的证明。根据级联估计量的形式（见 (III.6)）以及定理 IV.2 的第 2 项，我们有：当 $|U| \to \infty$ 时，$\widehat{\mathcal{C}}(T^{\alpha\beta})$ 以趋于 1 的概率包含在 $\mathcal{N}_{v^*}((\alpha+\beta)T^{\alpha\beta})$ 内。此外，还有

$$\begin{aligned}
\left|\widehat{\mathcal{C}}(T^{\alpha\beta})\right| &\overset{(a)}{\leq} \left|\mathcal{N}_{v^*}\left((\alpha+\beta)F_{v^*}^{\alpha}\left(\frac{15\log|U|}{p(1-2\epsilon)^2}\right)\right)\right|\\
&\overset{(b)}{\leq} \left(q\left|\mathcal{N}_{v^*}\left(\alpha F_v^{\alpha}\left(\frac{15\log|U|}{p(1-2\epsilon)^2}\right)\right)\right|\right)^{r(\alpha+\beta)/\alpha}\\
&\overset{(c)}{\leq} \left(qf_v^{\alpha}\left(F_v^{\alpha}\left(\frac{15\log|U|}{p(1-2\epsilon)^2}\right)\right)\right)^{r(\alpha+\beta)/\alpha}\\
&\overset{(d)}{=} \left(\frac{15q\log|U|}{p(1-2\epsilon)^2}\right)^{r(\alpha+\beta)/\alpha}.
\end{aligned}$$

上文中，(*a*) 由定理 IV.2 的第 1 项得出；(*b*) 归因于引理 D.5；(*c*) 成立是因为对于任意 $t \geq 0$，$f_v^{\alpha}(t) \geq \mathcal{N}_v(\alpha t)$，且 $v \in V(G)$；(*d*) 则利用了 F_v^{α} 是 f_v^{α} 的反函数这一事实。注意到当 $c > r(\alpha+\beta)/\alpha$ 时，最终的上界至多为 $\log^c|U|$，即可得出所需结论。 □

附录 D
基本图论结果

在本附录中，我们将证明一些简单但有用的图论结果。

引理 D.1. 假设 G 是连通的，且其最大度为 Δ。那么对于所有 $v \in V(G)$ 以及所有 $t \geq 0$，都有$|\mathcal{N}_v(t)| \leq 2\Delta^t$。

证明。令 $\partial\mathcal{N}_v(t)$ 表示与 v距离恰好 t 的顶点集合。于是我们有不等式 $|\partial\mathcal{N}_v(t)| \le \Delta|\partial\mathcal{N}_v(t-1)|$ 对于所有 $t \ge 1$ 均成立，这意味着 $|\partial\mathcal{N}_v(t)| \le \Delta^t$。接下来我们有

$$\begin{aligned}|\mathcal{N}_v(t)| &= \sum_{s=0}^{t} |\partial\mathcal{N}_v(s)| \le \sum_{s=0}^{t} \Delta^s \\ &= \frac{\Delta^{t+1}-1}{\Delta-1} \le \left(\frac{\Delta}{\Delta-1}\right)\Delta^t.\end{aligned}$$

由于 G 是连通的，因此 $\Delta \ge 2$。由此可得，$\Delta/(\Delta-1) \le 2$，这就证明了所需结论。 □

引理 D.2. 令 $\alpha > 0$ 和 $v \in V(G)$。对于任意 $1 \le t_1 \le t_2$，均有

$$f_v^\alpha(t_1) \le \frac{2t_1}{t_2} f_v^\alpha(t_2).$$

证明。由于 $f_v^\alpha(t)$ 是递增项之和，因此有$f_v^\alpha(t_1)/(t_1+1) \le f_v^\alpha(t_2)/(t_2+1)$。由此可得

$$f_v^\alpha(t_1) \le \frac{t_1+1}{t_2+1} f_v^\alpha(t_2).$$

由于 $t_1 + 1 \le 2t_1$ 且 $t_2 + 1 \ge t_2$，所需结论即可得证。 □

引理 D.3. 对于任意 $0 \le t_1 \le t_2$，均有 $f_v^\alpha(t_2) - f_v^\alpha(t_1) \ge t_2 - t_1$。

证明。由于对于所有 $s \ge 0$，$|\mathcal{N}_v(\alpha s)| \ge 1$，因此有

$$f_v^\alpha(t_2) - f_v^\alpha(t_1) = \sum_{s=t_1+1}^{t_2} |\mathcal{N}_v(\alpha s)| \ge t_2 - t_1.$$

□

引理 D.4. 当 $z \to \infty$，$F_v^\alpha(z) \sim F_v(\alpha z)/\alpha$。

证明。令 $\tilde{f}_v^\alpha(t)$ 和 $\tilde{f}_v(t)$ 分别表示 $f_v^\alpha(t)$ 和 $f_v(t)$ 的连续时间版本（对于 $t \in \mathbb{Z}_{\ge 0}$），它们由线性插值得到。为建立 f_v^α 与 f_v之间的联系，我们可以写成

$$\tilde{f}_v^\alpha(t) = \int_0^t |\mathcal{N}_v(\alpha s)| ds = \frac{1}{\alpha}\int_0^{\alpha t} |\mathcal{N}_v(\theta)| d\theta = \frac{1}{\alpha}\tilde{f}_v(\alpha t). \tag{D.1}$$

接下来，定义 $\tilde{F}_v^\alpha$和 $\tilde{F}_v$分别为 $\tilde{f}_v^\alpha$和 $\tilde{f}_v$的反函数。对于任意 $t \ge 0$，有

$$\frac{1}{\alpha}\tilde{F}_v(\alpha\tilde{f}_v^\alpha(t)) = \frac{1}{\alpha}\tilde{F}_v(\tilde{f}_v(\alpha t)) = t = \tilde{F}_v^\alpha(\tilde{f}_v^\alpha(t)),$$

其中，第一个等式由关系式 (D.1) 得出。由于 $\tilde{f}_v^\alpha$是连续函数，因此对于任意 $z \ge 0$，$\tilde{F}_v^\alpha(z) = F_v(\alpha z)/\alpha$。

最后，注意到 $\tilde{F}_v^\alpha$ 并且 F_v 是 F_v^α和 F_v的线性插值 $\tilde{F}_v$版本。此外，根据引理 D.3，F_v^α和 F_v至多线性增长。因此，当 $z \to \infty$时，$F_v^\alpha(z) \sim F_v^\alpha(z)$，且$F_v(z) \sim F_v(z)$，所需结论随即成立。 □

引理 D.5. 假设条件 (IV.1) 成立。那么，对于非负整数 t 和 k，任意 $v \in V(G)$ 均有$|\mathcal{N}_v(kt)| \le (q|\mathcal{N}_v(t)|^r)^k$。

证明。令 $k \ge 0$ 为一个整数。注意到 $\mathcal{N}_v((k+1)t) \subseteq \bigcup_{u\in\mathcal{N}_v(kt)} \mathcal{N}_u(t)$，有

$$|\mathcal{N}_v((k+1)t)| \le \sum_{u\in\mathcal{N}_v(kt)} |\mathcal{N}_u(t)| \le q|\mathcal{N}_v(t)|^r |\mathcal{N}_v(kt)|,$$

其中，第一个不等式由并集界得到，第二个不等式由 (IV.1) 得到。利用初始条件 $|\mathcal{N}_v(0)| = 1$ 求解该递推式，即可证明此引理。 □