刻画基于时效性的闲聊网络中的信息准确性

Emirhan Tekez，*IEEE*研究生会员，Melih Bastopcu，*IEEE*会员，Sinan Gezici，*IEEE*会士

arXiv:2603.02197v1 [cs.IT] 2 Mar 2026

摘要——**我们研究基于时效性的闲聊网络中的信息准确性，其中源按照具有 M 个状态的连续时间马尔可夫链（CTMC）演化，并向由 n 个节点组成的网络分发状态更新。除直接源更新外，节点通过信息传播交换其本地存储的数据包，并仅根据传入数据包是否比其本地副本更新来接受传入数据包。因此，一个节点可能拥有网络中最新的数据包，但仍不具有当前源状态。为量化在此类信息扩散方案下网络中流动的准确信息量，我们引入两个准确度指标：平均准确度，定义为任意给定子集中携带准确信息的节点的预期比例，以及基于新鲜度的准确度，定义为任意给定子集中最新节点的准确度。利用随机混合系统（SHS）框架，我们首先推导稳态平衡方程，并得到矩阵值递推，用以刻画二元连续时间马尔可夫链下全连接闲聊网络中的这些指标。随后，我们采用联合连续时间马尔可夫链方法，将分析扩展到一般多状态信息源。最后，我们量化因直接源推送与闲聊交换而信息准确的节点比例。我们通过数值分析验证研究结果，并给出渐近见解。**

索引词—**信息准确性、谣言传播网络、错误信息传播、基于时效性的信息传播**

I. 引言

现代网络化系统，例如自动驾驶车队[1], [2], 无线传感器网络 [3]–[5],和智能工厂，日益依赖快速变化的信息源。在此类场景中，快速传递更新至关重要，这推动了实时状态更新及信息年龄（AoI）指标相关文献的迅速发展 [2], [6]–[8]。AoI量化接收方收到的最新更新有多旧，但并不直接反映所接收内容是否准确。更具体地说，对于马尔可夫源，新鲜度和准确度可能彼此独立；源状态发生跃迁后，新接收的更新可能立即变得错误，而如果源的连续时间马尔可夫链（CTMC）回到相关状态，较旧的更新反而可能仍然正确。这促成了面向准确性的指标和建模方法，包括基于版本的马尔可夫信源监测视角 [9] 、错误信息年龄（AoII）指标 [10], 以及针对AoII类目标的策略设计 [11], [12]。

此外，信息扩散是一种经过充分研究的协议，用于点对点交换，无需集中协调 [13]–[16]。在信息年龄的研究背景下，及时谣言传播 已在不同拓扑结构和运行条件下得到研究 [17]–[27]。在本文中，该术语指一种基于时间戳的更新规则：只有当传入数据包比本地副本更新鲜时，节点才会接受它。与本文研究问题最相关的工作，是近期关于谣言传播网络中错误信息的研究成果，见 [24]和 [28]。[24] 的作者研究了不诚实节点和传输错误如何影响信息传播过程中真相的传播。在他们的模型中，真相被用作信息准确性的替代指标，但并未明确跟踪节点内容，而本文对此进行了跟踪。另一方面， [28]关注的是：当传输的数据包中的错误数量在网络中传播时按马尔可夫链演变，错误信息如何在谣言传播网络中扩散。然而，这些研究没有考虑源状态转移可能使信息变得准确，即使对于陈旧节点也是如此，例如源按照连续时间马尔可夫链演变的情况。

我们的工作还与针对马尔可夫源、不断发展的语义感知远程估计和监测研究方向相关。在该方向中，更新的价值取决于源状态的含义以及所产生的误差类型，而不仅仅取决于时效性 [29]–[32]。在相关文献中，通信决策通常被表述为直接源—监测器链路上的、可能依赖于状态的采样／查询控制问题，需要在更新成本与对马尔可夫状态作出错误推断的风险之间进行权衡 [29]–[32]。然而，这些研究与我们的工作存在差异，因为谣言传播网络涉及大量节点间的数据包交换，其决策基于局部新鲜度比较，而不是仅仅基于对源的直接观测。

受这些考量启发，我们研究由 M 状态连续时间马尔可夫链源驱动的基于时效性的流言网络中的信息准确度。我们聚焦于广泛研究的全连接（FC）流言网络 [21]–[24], [26], [33]。源将其当前状态推送给节点，而节点传播其存储的数据包，并且仅当传入数据包比本地副本更新鲜时才予以保留，这与基于版本的传播方式一致 [9], [23], [34]。由于节点仅依据时效性决定是否接受数据包，因此即使某个节点持有网络中最新鲜的数据包，该数据包仍可能与当前源状态不符。我们通过针对任意 k 节点子集定义的两项指标来量化这一效应：(i)平均准确度，即子集中持有准确信息的节点所占比例的期望值；以及 (ii) 基于新鲜度的准确度，即子集中最新鲜节点的准确度。

我们的贡献可列举如下：（i）我们使用随机混合系统（SHS）方法，对准确信息和不准确信息如何

作者来自比尔肯特大学电气与电子工程系，安卡拉，T¨urkiye（电子邮箱：emirhan.tekez@bilkent.edu.tr；bastopcu@bilkent.edu.tr；gezici@ee.bilkent.edu.tr）。本研究得到 TUBITAK 2232-B 项目（项目编号：124C533）和 TUBITAK 2210-A 项目的支持。

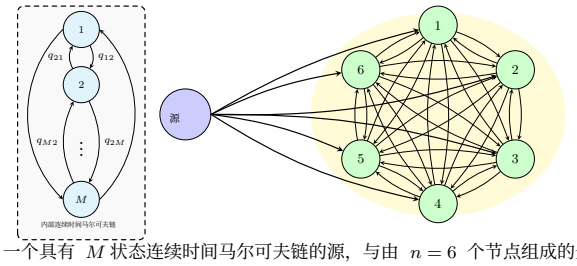


图 1. 一个具有 M 状态连续时间马尔可夫链的源，与由 $n = 6$ 个节点组成的全连接信息传播网络相连。

在基于时效性的闲聊网络中传播进行了定量分析 [35], [36]。（ii）我们建立了矩阵值递推方程，用于量化二元马尔可夫信源下网络中持有准确信息的节点比例。（iii）我们采用联合连续时间马尔可夫链方法，将研究结果扩展到具有 M 个状态的有限 CTMC。（iv）我们量化了因 < 杰 style id='6'>源推送 而保持准确的节点比例，以及因<杰 style id='8'>信息源反转而保持准确的节点比例。值得注意的是，这是首项在明确考虑节点信息内容、且不使用代理变量简化分析的情况下，完整刻画基于时效性的闲聊网络中信息准确性的研究。（v）我们深入分析了渐近机制下闲聊网络中信息准确性的表现。我们表明，在基于时效性的闲聊网络中，提高点对点 Gossip 速率对改善准确度的效果，远不及提高源推送速率。（vi）此外，我们证明，网络中持有任意给定信息的节点数量仅取决于源节点连续时间马尔可夫链的平稳分布，而与任何传输速率无关。另外，我们通过不同源推送速率和 Gossip 速率下的仿真结果，验证了我们的发现。

II. 系统模型与问题表述

我们考虑一个由 n 个节点组成的全连接（FC）信息传播网络，$\mathcal{N} = \{1, \ldots, n\}$，以及一个用于传播其最新状态更新的源。源处的信息按照一个不可约的有限状态连续时间马尔可夫链（CTMC）演化，该链共有 M 个状态，即 $Q(t) \in \{1, \ldots, M\}$，并按照已知的生成矩阵 $\mathbf{Q}$ 自行更新，其元素标记为 q_{ij}，表示从状态 i到状态 j的跃迁。源以速率为 $\frac{\lambda_s}{n}$ 的稀疏泊松过程向网络中的每个节点 $i \in \mathcal{N}$发送更新（即 CTMC 的当前状态）。我们假设源始终持有新鲜且真实的信息。图 1 展示了一个由 $n = 6$ 个节点组成的相关系统简单示例。

当连续时间马尔可夫链的状态发生变化时，源会生成一个新的更新版本。我们将时刻 t 源处最新的信息版本记为$V_0(t)$。因此，每当生成一个新的更新版本时，源处的信息版本 $V_0(t)$就会加一。类似地，我们将时刻 t节点 i 处最新的更新版本记为 $V_i(t)$，其内容记为 $S_i(t)$。当源在时刻t向节点 i发送信息时，节点 i 会同时接收到连续时间马尔可夫链的当前状态以及以 $\{V_0(t)\ Q(t)\}$形式表示的版本信息。节点希望尽可能准确地跟踪连续时间马尔可夫链的当前状态。为此，节点始终接受来自源的更新，因为源始终维护着最新且准确的信息。因此，如果节点 i 在时刻 t收到来自源的更新，该节点会将其信息更新为 $V_i(t) = V_0(t)$ 和 $S_i(t) = Q(t)$。为了衡量节点处的信息新鲜度，我们采用一种称为信息版本年龄（VAoI）的指标，该指标在 [21]‚[34]中提出。时刻 t 节点 i 的信息版本年龄记为 $X_i(t)$，其定义为

$$X_i(t) = V_0(t) - V_i(t). \tag{1}$$

由于源的更新率 λ_s 有限，且由各节点均分，网络规模增大可能导致源更新不足以使所有节点都保持准确信息。为了提高信息新鲜度，节点还通过流言协议 [17], [21]共享最新信息。为此，网络中的每个节点都根据另一个稀疏泊松过程向其他节点发送更新，其中节点间流言传播以速率 $\frac{\lambda}{n-1}$ 进行。当两个节点相互传播信息时，由于它们无法判断传入数据包是否准确，因此只考虑传入数据包的版本年龄。更正式地说，当节点 i 尝试向节点 j发送数据包时，节点 j 处的版本年龄按照以下规则演变：

$$X_j(t) = \min\{X_i(t^-), X_j(t^-)\}, \tag{2}$$

其中 t^- 表示信息交换之前的时刻。因此，当节点 i 向节点j发送其信息时，节点 j 会检查传入信息的版本；如果传入信息更新，即 $V_i(t) > V_j(t)$，节点 j 便接受节点 i 的信息；否则，它将继续保留当前信息。因此，节点 j 处的版本年龄如式 (2) 所示。为了分析平均版本年龄，我们需要跟踪任意一个包含 k 个元素的子集的版本年龄，该子集记为 X_{A_k}，其定义如下：

$$X_{A_k}(t) \triangleq \min_{j \in A_k} X_j(t). \tag{3}$$

换言之，集合的版本年龄由集合 A_k 中最新鲜的信息决定。

在这项工作中，我们旨在刻画网络中节点的平均信息准确度。为此，我们将节点 i 的准确度定义为 $C_i(t)$：

$$C_i(t) = \begin{cases} 1, & S_i(t) = Q(t), \\ 0, & S_i(t) \neq Q(t). \end{cases} \tag{4}$$

根据定义 (4)，准确度是一种二元指标：当节点 i 处的信息与源处的信息相同时，取值为 1；否则取值为 0。这种准确度概念曾以二元新鲜度之名在 [37], [38], 中进行研究，但这些工作仅涉及单个源监视器，且未考虑该指标在包含大量点对点交换的流言传播网络中的演变情况。同样，我们将 $C_{A_k}(t)$ 定义为网络中任意子集的平均准确度，该子集包含 k 个节点，记为 A_k，其中 C_A t 定义为

$$C_{A_k}(t) = \frac{C_1(t) + C_2(t) + \cdots + C_k(t)}{k} = \frac{1}{k} \sum_{\ell \in A_k} C_\ell. \tag{5}$$

$_k$( )

请注意，准确度的定义仅取决于源和相关节点的内容，而不取决于

节点的版本年龄。因此，一般来说，节点可以通过两种方式获得准确信息：i）节点获取源处的最新信息，因此其版本年龄为 0，并且拥有准确信息；ii）节点可能拥有较旧版本的信息，但如果源返回到相关状态，这个陈旧数据包仍可能是准确的。换言之，携带较旧版本信息的过时节点可能拥有正确的信息，这正是信息及时性与信息准确性之间的区别。我们还注意到，我们在 (4) 中定义的信息准确性度量是 [10] 中所称的更一般度量——信息不准确年龄（AoII）——的一种特例。最后，正如前文所述，在节点之间进行信息扩散时，节点并不知道自己持有的信息是否准确。因此，信息交换仍然依据 (2) 中描述的信息新鲜度进行。于是，当节点 i 向节点 j 发送状态更新时，节点 j 处的信息准确性变化如下：

$$C_j(t) = \begin{cases} C_i(t^-), & X_i(t^-) \le X_j(t^-), \\ C_j(t^-), & X_i(t^-) > X_j(t^-). \end{cases} \tag{6}$$

换言之，如果节点 i携带更新鲜的信息，则节点j的准确度会更新为与节点 i一致；否则保持不变。

在以下各节中，我们分析网络中各节点在二元信息源（即 $M = 2$）以及具有一般状态数的源（即 $M > 2$）下的信息准确度。

III. 二元信息源

我们首先研究二元连续时间马尔可夫链信息源模型（即 $M = 2$）下的平均准确度，其生成矩阵为

$$\mathbf{Q} = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}, \tag{7}$$

其中 $q_{11} = -q_{12}$ 且 $q_{22} = -q_{21}$。换言之，信息源以速率 q_{12}从状态1转移到状态2，并以速率 q_{21}从状态2转移到状态1。该二进制连续时间马尔可夫链的平稳分布为

$$\pi_1 = \frac{q_{21}}{q_{12} + q_{21}}, \qquad \pi_2 = \frac{q_{12}}{q_{12} + q_{21}}. \tag{8}$$

为了评估持有正确信息的节点的期望数量，即 $c_k = \lim_{t\to\infty} \mathbb{E}[C_{A_k}(t)]$, 我们使用随机混杂系统（SHS）框架 [35], [36]对该系统进行建模。混杂状态由下式给出

$$\big(Q(t), \mathbf{Z}(t)\big) = \big(Q(t), \mathbf{C}(t), \mathbf{X}(t)\big), \tag{9}$$

其中，离散模式 $Q(t) \in \{1, 2\}$ 表示源的连续时间马尔可夫链状态，连续分量由 $\mathbf{C}(t) = [C_1(t) \dots, C_n(t)] \in \{0, 1\}^n \subset \mathbb{R}^n$ 构成，其中 $C_i(t)$ 表示节点 i 在时刻 t 的准确度，而 $\mathbf{X}(t) = [X_1(t) \dots, X_n(t)] \in \mathbb{R}^n$ 表示网络中所有节点的时龄向量。将它们合并为 $\mathbf{Z}(t) = (\mathbf{C}(t), \mathbf{X}(t)) \in \mathbb{R}^{2n}$。在跳变之间，连续状态保持不变，即它遵循微分方程 $\dot{(\mathbf{C}(t), \mathbf{X}(t))} = \mathbf{0}_{2n}$。

该随机混合系统的可能跃迁集合为

$$\mathcal{E} = \{(0,0,1\to2), (0,0,2\to1)\} \cup \{(0, j) : j \in \mathcal{N}\} \cup \{(i, j) : i, j \in \mathcal{N}, i \ne j\}. \tag{10}$$

其中，节点 0 表示源。转移 $(0, 0, 1 \to 2)$ 和 $(0, 0, 2 \to 1)$ 分别表示源状态从 1 翻转到 2 和从状态 2 翻转到 1。转移 $(0, j)$ 表示源向节点 j 发送更新的情况，而转移 (i, j) 表示从节点 i 到节点 j 的信息扩散事件。请注意，在本研究中，我们假设闲谈事件是有向的；对于每个满足 $i \ne j$ 的有序对 (i, j)，节点 i 向节点 j 发送一个数据包，并且在所描述的信息扩散事件中，只有节点 j 的准确度和版本年龄会发生变化。每次转移 $e \in \mathcal{E}$ 都会根据

$$\big(Q(t^+), \mathbf{Z}(t^+)\big) = \phi_e\big(Q(t^-), \mathbf{Z}(t^-)\big), \tag{11}$$

为简洁起见，下面仅给出 (13)– (19) 中的$\mathbf{Z}$分量。对于 (10) 中指定的跃迁，依赖于模式的速率映射 $\lambda_e(q)$ 定义为

$$\lambda_e(q) = \begin{cases} q_{12}\,\mathbb{1}_{\{q=1\}}, & e = (0,0,1\to2), \\ q_{21}\,\mathbb{1}_{\{q=2\}}, & e = (0,0,2\to1), \\ \dfrac{\lambda_s}{n}, & e = (0, j),\ j \in \mathcal{N}, \\ \dfrac{\lambda}{n-1}, & e = (i, j),\ i, j \in \mathcal{N},\ i \ne j, \end{cases} \tag{12}$$

其中，$\mathbb{1}_{\{x\}}$ 是指示函数，仅当x 为真时取值为 1，否则取值为 0。由 (12) 可知，转移 $(0, 0, 1 \to 2)$ 仅在模式 $Q = 1$ 下以速率q_{12}激活，而转移 $(0, 0, 2 \to 1)$ 仅在模式 $Q = 2$下以速率 q_{21} 激活；推送速率和信息传播速率则不依赖于源模式。

对于每个节点 $\ell \in \mathcal{N}$，状态 $(\mathbf{C}(t), \mathbf{X}(t))$ 按照 (10) 中指定的转移，根据以下重置映射演化。首先，我们给出单个节点 $\ell \in \mathcal{N}$ 的版本年龄重置映射：

$$X'_\ell = \begin{cases} X_\ell + 1, & e \in \{(0,0,1\to2), (0,0,2\to1)\}, \ell \in \mathcal{N}, \\ 0, & e = (0, j),\ \ell = j \in \mathcal{N}, \\ \min\big(X_i, X_j\big), & e = (i, j),\ i, j \in \mathcal{N},\ i \ne j,\ \ell = j, \\ X_\ell, & \text{otherwise.} \end{cases} \tag{13}$$

对于任何源反转事件，由于源在每次状态转移时都会生成一个新鲜数据包，因此每个节点的信息版本年龄都会增加 1。在信息传播事件中，只有当传入数据包更新鲜时，接收方才会采用发送方的版本年龄，如 (2) 所示。

同样，节点 $\ell \in \mathcal{N}$ 的准确性指示变量演化如下：

$$C'_\ell = \begin{cases} 1 - C_\ell, & e \in \{(0,0,1\to2), (0,0,2\to1)\}, \ell \in \mathcal{N}, \\ 1, & e = (0, j),\ \ell = j \in \mathcal{N}, \\ C_{\arg\min_{\ell \in \{i,j\}} X_\ell}, & e = (i, j),\ i, j \in \mathcal{N},\ i \ne j,\ \ell = j, \\ C_\ell, & \text{otherwise.} \end{cases} \tag{14}$$

每次发生信息源反转时，任意节点的准确度都会按照式（14）中的第一种情况进行反转。由于我们假设源始终持有真实且新鲜的信息，因此在接收到来自源的信息后，X_ℓ 和 C_ℓ 分别设为 0 和 1。式（14）中的第三种情况对应于节点 i 向节点 j 发送更新的情形。如前所述，由于节点无法获知其信息的准确度，节点 j 会根据信息的新鲜度决定接受还是拒绝节点 i 的更新。因此，根据式（14）中的第三种情况，节点 j 的准确度将变为包含节点 i 和 j 的集合中最新鲜信息的准确度。

我们可以将式 (13) 和 (14) 中针对单个节点的重置映射扩展到网络中包含 k 个节点的任意子集，以便

研究第二节中描述的 C_{A_k} 和 X_{A_k} 量。首先，我们为任意集合 A_k 提供如下版本年龄重置映射：

$$X'_{A_k}=\begin{cases} X_{A_k}+1, & e\in\{(0,0,1\to 2),\,(0,0,2\to 1)\},\\ 0, & e=(0,j),\ j\in A_k,\\ X_{A_{k+1}}, & e=(i,j),\ i\in\mathcal{N}\setminus A_k,\ j\in A_k,\\ X_{A_k}, & \text{otherwise.}\end{cases} \tag{15}$$

我们注意到，集合 A_k 的版本年龄，即 $X_{A_k}(t)$，定义见 (3)。当源处的信息更新时，集合 A_k 中最新节点的版本年龄，也就是集合 A_k 的版本年龄，会增加 1，这对应于 (15) 中给出的第一种情况。每当源向集合 A_k 中的某个节点发送更新时，接收节点 j 就会成为集合 A_k 中最新的节点，从而将该集合的版本年龄重置为 0。如果集合 A_k 外的节点 i 向集合 A_k 中的某个节点发送更新，那么若传入数据包更新鲜，接收节点就会复制外部节点的版本年龄。因此，我们需要考察扩展集合 $A_k\cup\{i\}$ 的最小年龄，以确定跃迁后的值。由于网络是全连接的，根据对称性，我们可以得出 $X_{A_k\cup\{i\}}=X_{A_{k+1}}$。类似地，任意集合 A_k 的准确度重置映射如下：

$$C'_{A_k}=\begin{cases} 1-C_{A_k}, & e\in\{(0,0,1\to 2),(0,0,2\to 1)\},\\ C_{A_k}+\dfrac{1-C_j}{k}, & e=(0,j),\ j\in A_k,\\ C_{A_k}+\dfrac{F_{\{i,j\}}-C_j}{k}, & e=(i,j),i,j\in\mathcal{N},\ i\neq j,\ j\in A_k,\\ C_{A_k}. & \text{otherwise.}\end{cases} \tag{16}$$

每当源处的信息发生变化时，集合 A_k 的准确度就会按照 (16) 的第一种情况翻转。当源向集合 A_k 中的节点 j 发送更新时，如果节点 j的准确度为 0，则准确度增加 $\frac{1}{k}$。类似地，如果不在集合 A_k 中的节点 i 向集合 A_k 中的节点 j 发送更新，那么当节点 i 携带着更新鲜的更新时，节点 j 会接受该更新。在这种情况下，当节点 i 携带准确信息而节点 j 携带不准确信息时，集合 A_k *i*) 的准确度增加 $\frac{1}{k}$；如果两个节点的准确度相同，则 *ii*) 保持不变；如果节点 i 携带不准确信息而节点 j 携带准确信息，则 *iii*) 减少 $\frac{1}{k}$。

第三种情况中的此类操作要求我们定义一个新函数，记为 F_{A_k}，表示任意集合A_k 中最新节点的准确度，其定义如下

$$F_{A_k}(\mathbf{C},\mathbf{X})\triangleq C_{\arg\min_{\ell\in A_k}X_\ell}. \tag{17}$$

这对应于如下形式的测试函数

$$\psi_k(Q,\mathbf{Z})=F_{A_k}(\mathbf{C},\mathbf{X}), \tag{18}$$

该函数不显式依赖离散模式 Q，且在跳变之间保持恒定。接下来，我们定义重置映射，它描述了 F_{A_k} 在 (10) 中的转移下如何变化，即 $F'_{A_k}=\psi_k\big(\phi_e(Q,\ \mathbf{Z})\big)$:

$$F'_{A_k}=\begin{cases} 1-F_{A_k}, & e\in\{(0,0,1\to 2),\,(0,0,2\to 1)\},\\ 1, & e=(0,j),\ j\in A_k,\\ F_{A_{k+1}}, & e=(i,j),\ i\in\mathcal{N}\setminus A_k,\ j\in A_k,\\ F_{A_k}, & \text{otherwise.}\end{cases} \tag{19}$$

当源中的信息发生更新时，F_{A_k} 会翻转，这与（16）中定义的第一种情况类似。当源向集合 F_{A_k} 中的某个节点发送更新时，接收方节点会成为该集合中的最新节点，并继承源的准确度，因此有 $F'_{A_k}=1$。在信息传播事件中，当 A_k 外部的节点向内部的某个节点进行传播时，A_k 基于新鲜度的准确度将与扩展集合 $A_k\cup\{i\}$ 的准确度一致，并且根据对称性，等于 $F_{A_{k+1}}$。

回顾随机混合系统稳态平衡恒等式。对于混杂状态的任意（时不变）测试函数 ψ，有

$$\frac{d}{dt}\,\mathbb{E}\big[\psi\big(Q(t),\mathbf{Z}(t)\big)\big]=0. \tag{20a}$$

因为在稳态条件下，测试函数的期望值不会随时间变化。扩展生成元 $\mathcal{L}\psi(Q,\ \mathbf{Z})$ 通过对所有跃迁类型 $e\in\mathcal{E}$ 的贡献求和，给出 $\psi(Q(t),\ \mathbf{Z}(t))$ 的瞬时变化：

$$\mathcal{L}\psi(Q,\mathbf{Z})=\sum_{e\in\mathcal{E}}\lambda_e(Q)\Big[\psi\big(\phi_e(Q,\mathbf{Z})\big)-\psi(Q,\mathbf{Z})\Big]. \tag{20b}$$

利用稳态下期望变化为零这一事实，可得

$$\sum_{e\in\mathcal{E}}\mathbb{E}\Big[\lambda_e\big(Q(t)\big)\Big(\psi\big(\phi_e(Q(t),\mathbf{Z}(t))\big)-\psi\big(Q(t),\mathbf{Z}(t)\big)\Big)\Big]=0. \tag{20c}$$

对于下文介绍的特定 $\psi(\cdot)$ 选择（例如，F_{A_k} 及其带模式标签的变体），(20c) 可得随机混合系统平衡方程。

A. 带模式标签的新鲜度准确性刻画

由于源推送事件的发生率取决于跃迁 e 之前系统所处的离散模式，我们首先推导基于新鲜度的准确度 F_{A_k} 的精确刻画，并明确追踪源模式 $Q(t)$。对于 $m\in\{1,2\}$ 和 $k\in\{1,\dots,n\}$，我们将带模式标签的期望定义为

$$f_k^{(m)}\triangleq\lim_{t\to\infty}\mathbb{E}\big[F_{A_k}(t)\,\mathbb{1}_{\{Q(t)=m\}}\big], \tag{21}$$

以及 $\mathbf{f}_k\triangleq[f_k^{(1)},f_k^{(2)}]^T$。其中，$[\cdot]^T$ 表示向量的转置。这两个量之和给出了无条件的基于新鲜度的准确度，如下所示：

$$f_k\triangleq f_k^{(1)}+f_k^{(2)}=\lim_{t\to\infty}\mathbb{E}[F_{A_k}(t)]. \tag{22}$$

为得到平衡方程，我们使用测试函数

$$\psi_k^{(1)}(Q,\mathbf{Z})=F_{A_k}(\mathbf{C},\mathbf{X})\,\mathbb{1}_{\{Q=1\}}, \tag{23}$$

$$\psi_k^{(2)}(Q,\mathbf{Z})=F_{A_k}(\mathbf{C},\mathbf{X})\,\mathbb{1}_{\{Q=2\}}. \tag{24}$$

其中，$F_{A_k}(\mathbf{C},\mathbf{X})$ 表示集合 A_k 中最新节点的准确度，该集合定义于式（17）中，并且我们有 $f_k^{(m)}=\lim_{t\to\infty}\mathbb{E}[\psi_k^{(m)}]$。为方便记号，我们定义以下速率

$$\alpha_k\triangleq\frac{k\lambda_s}{n},\quad \beta_k\triangleq\frac{k(n-k)\lambda}{n-1},\quad \gamma_k\triangleq\frac{k(k-1)\lambda}{n-1}, \tag{25}$$

where分别是集合 A_k 中节点的总源推送速率、集合 A_k 外部节点向集合 A_k 内部节点传播信息的速率，以及集合 A_k 内部节点之间传播信息的速率。注意，α_k、β_k 和 γ_k 均不依赖于离散模式 Q。

考虑式（23）中给出的测试函数 $\psi_k^{(1)}(Q,\mathbf{Z})=F_{A_k}\mathbb{1}_{\{Q(t)=1\}}$。有三类事件会引起变化：*a*) 信息源反转、*b*) 源推送，以及 *c*) 信息传播事件。因此，$\psi_k^{(1)}(Q,\mathbf{Z})$ 的随机混合系统平衡方程可写为：

$$\lim_{t\to\infty}\mathbb{E}\Big[\Big(-q_{12}F_{A_k}+\alpha_k(1-F_{A_k})+\beta_k(F_{A_{k+1}}-F_{A_k})\Big)\cdot\mathbb{1}_{\{Q=1\}}+\Big(q_{21}(1-F_{A_k})\Big)\mathbb{1}_{\{Q=2\}}\Big]=0. \tag{26}$$

式（26）中的第一项源于信息源反转跃迁 $(0,0,1 \to 2)$，其发生率为 $q_{12}\mathbb{1}_{\{Q(t)=1\}}$。该跃迁发生后，源状态变为 $Q'(t) = 2$。根据式（19），状态 1 的基于新鲜度的准确度变为 0（因为源现在处于状态 2），因此，该跃迁引起的总体变化可写为 $(0-F_{A_k})\mathbb{1}_{\{Q=1\}}$。类似地，式（26）中的第二项对应于跃迁 $(0,0,2 \to 1)$，其发生率为 $q_{21}\mathbb{1}_{\{Q(t)=2\}}$。在这种情况下，跃迁后新的源状态变为 $Q'(t) = 1$，而式（19）表明状态 1 的基于新鲜度的准确度变为 $1 - F_{A_k}$。因此，该跃迁引起的总体变化可表示为 $(1-F_{A_k})\mathbb{1}_{\{Q=2\}}$。式（26）中的第三项表示源更新集合 A_k 中的一个节点，此时 F_{A_k} 取值为 1。该跃迁引起的总体变化可写为 $(1-F_{A_k})\mathbb{1}_{\{Q=1\}}$。最后，式（26）中的最后一项表示集合 A_k 外的一个节点向集合 A_k 内的一个节点传播信息的情况，其发生率为更新率 β_k。此次更新后，集合 A_k 的准确度变为超集 $F_{A_{k+1}}$ 的基于新鲜度的准确度，该超集包含发送节点。因此，该跃迁引起的变化可写为 $(F_{A_{k+1}} - F_{A_k})\mathbb{1}_{\{Q=1\}}$。

在对(26)取期望后，我们得到

$$-q_{12}f_k^{(1)}+q_{21}(\pi_2-f_k^{(2)})+\alpha_k(\pi_1-f_k^{(1)})+\beta_k(f_{k+1}^{(1)}-f_k^{(1)}) = 0, \tag{27}$$

对于 $k = 1, \ldots\ldots, n-1$。对$\psi_k^{(2)}(Q, \mathbf{Z})= F_{A_k}\mathbb{1}_{\{Q(t)=2\}}$重复类似分析后，我们可以得到：

$$q_{12}(\pi_1-f_k^{(1)})-q_{21}f_k^{(2)}+\alpha_k(\pi_2-f_k^{(2)})+\beta_k(f_{k+1}^{(2)}-f_k^{(2)}) = 0, \tag{28}$$

对于 $k = 1, \ldots\ldots, n-1$。结合(27)和(28)，我们得到如下形式的线性递推关系

$$W_k\mathbf{f}_k = \mathbf{v}_k + V_k\mathbf{f}_{k+1}, \qquad k = 1, \ldots, n-1, \tag{29}$$

with

$$W_k \triangleq \begin{bmatrix} q_{12}+\alpha_k+\beta_k & q_{21} \\ q_{12} & q_{21}+\alpha_k+\beta_k \end{bmatrix},\ V_k \triangleq \beta_k I_2,\ \mathbf{v}_k \triangleq \begin{bmatrix} q_{21}\pi_2+\alpha_k\pi_1 \\ q_{12}\pi_1+\alpha_k\pi_2 \end{bmatrix}, \tag{30}$$

其中 I_2 表示一个 2×2 单位矩阵。由于 W_k 对于 α_k, $\beta_k > 0$是可逆矩阵，我们将$\mathbf{f}_k$ 表示为

$$\mathbf{f}_k = W_k^{-1}\big(\mathbf{v}_k + V_k\mathbf{f}_{k+1}\big), \qquad k = 1, \ldots, n-1. \tag{31}$$

我们注意到，(31)意味着基于新鲜度的模式标记准确度的矩阵值反向递推。

当 $k = n$时，集合 A_n 包含所有节点。因此， $\beta_n = 0$且 $\alpha_n = \lambda_s$。平衡方程(27)–(28)简化为

$$q_{21}(\pi_2 - f_n^{(2)}) - q_{12}f_n^{(1)} + \lambda_s(\pi_1 - f_n^{(1)}) =0, \tag{32}$$

$$q_{12}(\pi_1 - f_n^{(1)}) - q_{21}f_n^{(2)} + \lambda_s(\pi_2 - f_n^{(2)}) =0. \tag{33}$$

联立求解 (32)–(33) 中的 2×2 系统以及 (8) 中的 (π_1, π_2)，可得

$$\mathbf{f}_n \triangleq \begin{bmatrix} f_n^{(1)} \\ f_n^{(2)} \end{bmatrix} = \frac{1}{(q_{12}+q_{21})(\lambda_s+q_{12}+q_{21})} \begin{bmatrix} q_{21}(\lambda_s+q_{21}) \\ q_{12}(\lambda_s+q_{12}) \end{bmatrix}. \tag{34}$$

给定 $\mathbf{f}_n$，式 (31) 中的递推关系确定了所有 $k = n-1, \cdots, 1$ 对应的 $\mathbf{f}_k$。完成了新鲜度型准确度的刻画后（这是分析平均准确度所必需的），我们现在继续研究网络的平均准确度。

B. 模式标记平均准确度的刻画

在式 (16) 中，我们展示了集合 A_k 的准确度（记为 C_{A_k}）如何根据网络中可能发生的跃迁而重置。我们可以看到，C_{A_k} 取决于集合 A_k 中最新节点的准确度（记为 F_{A_k}）。接下来，我们将明确刻画模式标记平均准确度，以完成分析，并推导相关表达式，从而深入了解在渐近机制下准确信息和不准确信息的传播方式。

定理 1. 令 $\mu \triangleq \lambda_s/n$ 表示每节点源推送速率，并令 $\mathbf{c} \triangleq [c^{(1)}, c^{(2)}]^T$ 表示由模式标记平均准确率组成的向量。那么，$\mathbf{c}$是以下线性方程组的唯一解：

$$\bar{W}\,\mathbf{c} = \mathbf{b} + \lambda\,\mathbf{f}_2, \tag{35}$$

其中

$$\bar{W} \triangleq \begin{bmatrix} q_{12}+\mu+\lambda & q_{21} \\ q_{12} & q_{21}+\mu+\lambda \end{bmatrix},\ \mathbf{b} \triangleq \begin{bmatrix} q_{21}\pi_2+\mu\pi_1 \\ q_{12}\pi_1+\mu\pi_2 \end{bmatrix},$$

并且 $\mathbf{f}_2 \triangleq \begin{bmatrix} f^{(1)}{}_2, f_2^{(2)} \end{bmatrix}$.因此，网络中任意基数为 $k \in \{1, 2, \cdots, n\}$ 的子集的无条件平均准确度为

$$c_k = c\ =\ \mathbf{1}^T\mathbf{c}, \qquad \mathbf{1} = [1\ \ 1]^T.$$

证明。对于 $m \in \{1, 2\}$ 和 $k \geq 1$，我们首先定义集合 A_k 的准确度在给定状态 $Q(t) =m$ 时的期望值为

$$c_k^{(m)} \triangleq \lim_{t\to\infty} \mathbb{E}\big[C_{A_k}(t)\,\mathbb{1}_{\{Q(t)=m\}}\big]. \tag{36}$$

接下来，我们将总体期望准确度定义为

$$c_k = c_k^{(1)} + c_k^{(2)}, \tag{37}$$

对所有 k均成立。与对模式标记的新鲜度基础准确度的分析类似，我们使用以下两个测试函数来计算平均准确度：

$$\hat{\psi}_k^{(1)}(Q, \mathbf{Z}) = C_{A_k}\,\mathbb{1}_{\{Q=1\}},\ \ \hat{\psi}_k^{(2)}(Q, \mathbf{Z}) = C_{A_k}\,\mathbb{1}_{\{Q=2\}}, \tag{38}$$

因此，我们有 $c_k^{(m)} =\lim_{t\to\infty}\mathbb{E}[\hat{\psi}_k^{(m)}]$。接下来，仿照 $F_{(1)A_k}$ 和 $F_{A_k}^{(2)}$ 的平衡方程，我们可以写出以下关于 $\hat{\psi}_k^{(1)}(Q, \mathbf{Z}) =C_{A_k}\,\mathbb{1}_{\{Q=1\}}$ 的平衡方程：

$$\lim_{t\to\infty}\mathbb{E}\Big[\Big(-q_{12}C_{A_k}+\alpha_k\Big(\tfrac{1-C_j}{k}\Big)+(\beta_k+\gamma_k)\Big(\tfrac{F_{\{i,j\}}-C_j}{k}\Big)\Big) \cdot \mathbb{1}_{\{Q=1\}} + \Big(q_{21}(1-C_{A_k})\Big)\mathbb{1}_{\{Q=2\}}\Big] = 0. \tag{39}$$

取期望并利用 $\frac{d}{dt}\,\mathbb{E}\left[\,\hat{\psi}\big(Q, \mathbf{Z}\big)\right] = 0$ 这一事实后，我们得到

$$q_{21}(\pi_2-c_k^{(2)})-q_{12}c_k^{(1)}+\tfrac{\lambda_s}{n}(\pi_1-c_1^{(1)})+\lambda(f_2^{(1)}-c_1^{(1)})=0 \tag{40}$$

.通过对 $\hat{\psi}_k^{(2)}(Q, \mathbf{Z})=C_{A_k}\mathbb{1}_{\{Q=2\}}$ 遵循类似步骤，我们得到$q_{12}(\pi_1-c_k^{(1)})-q_{21}c_k^{(2)}+\tfrac{\lambda_s}{n}(\pi_2-c_1^{(2)})+\lambda(f_{(2}{}_2-c_1^{(2)})) =0.$ 我们注意到，$C_{A_k}(t) =\frac{1}{k}\sum_{\ell=1}^{k} C_\ell(t)$ 如 (5) 中所给出。由于该网络是全连接且对称的，网络中的每个节点都具有相同的平均准确度。因此，所有 $c_k^{(m)}$ （对于 $k \geq 1$ ）彼此相等，即

是， $c_k^{(m)} = c^{(m)}$, 对所有 $k \in \{1, \ldots, n\}$, 以及 $m \in \{1, 2\}$。于是，(40) 和 (41) 中的准确度方程简化为

$$q_{21}(\pi_2 - c^{(2)}) - q_{12}c^{(1)} + \frac{\lambda_s}{n}\left(\pi_1 - c^{(1)}\right) + \lambda\left(f_2^{(1)} - c^{(1)}\right) = 0, \quad (42)$$

$$q_{12}(\pi_1 - c^{(1)}) - q_{21}c^{(2)} + \frac{\lambda_s}{n}\left(\pi_2 - c^{(2)}\right) + \lambda\left(f_2^{(2)} - c^{(2)}\right) = 0. \quad (43)$$

定义模式标记平均准确度向量 $\mathbf{c} \triangleq [c^{(1)}, c^{(2)}]^T$。于是，(42)-(43) 可简洁地写成式 (35) 所示的形式。

因此，在二进制连续时间马尔可夫链下，单个节点（等价地，对于任意 $k \in \{1, \cdots, n\}$ 的集合 A_k）的平均准确度为 $c \triangleq c^{(1)} + c^{(2)} = \mathbf{1}^T\mathbf{c}$；由对称性可得，对于所有 $k \geq 1$，均有 $c_k = c$。证明完毕。 □

接下来，我们研究几个可由定理 1 直接推出的极限，首先考虑源能够不受速率限制地向网络推送新信息的情形。

推论 1。固定 $\lambda \geq 0$，并令 $\mu = \lambda_s/n$。设 $\mathbf{c} = [c^{(1)}, c^{(2)}]^T$ 是 (35) 的唯一解。那么，当 λ_s （或等价地，μ）变得很大时，有

$$\lim_{\lambda_s \to \infty} c^{(m)} = \pi_m, \; m \in \{1, 2\}, \quad (44)$$

因此，极限 $\lim_{\lambda_s \to \infty} c = \pi_1 + \pi_2 = 1$. 这里， $\boldsymbol{\pi} \triangleq [\pi_1, \pi_2]^T$ 是源节点连续时间马尔可夫链的平稳分布。

证明。根据定理 1，$W = \mu I_2 + \tilde{W}$，其中 $\tilde{W} \triangleq \begin{bmatrix} q_{12} + \lambda & q_{21} \\ q_{12} & q_{21} + \lambda \end{bmatrix}$。若将两边除以 μ，则得到

$$\left(I_2 + \frac{1}{\mu}\tilde{W}\right)\mathbf{c} = \boldsymbol{\pi} + \frac{\lambda}{\mu}\mathbf{f}_2,$$

由于 $\mathbf{b}/\mu$ 随着 μ 任意增大而趋近于 $\boldsymbol{\pi}$，因此有 $\lim_{\mu \to \infty} \mathbf{c} = \boldsymbol{\pi}$，进而有 $\lim_{\mu \to \infty} c = 1$。 □

推论 1 直接源于以下假设：源始终持有新鲜且准确的信息。随着源推送速率任意增大，这些新鲜且准确的信息将压过所有闲谈事件，从而使网络中的所有节点都变得准确。

在下面的推论中，我们探讨当源停止向网络发送更新时，网络的平均准确度会如何变化。

推论 2。固定 $\lambda > 0$ 并令 $\mu = \lambda_s/n$。那么，当 λ_s （或等价地，μ）趋近于 *0* 时，模态标记准确度向量 $\mathbf{c}$ 收敛到 $\mathbf{c}_0 \triangleq$ 极限 $\lambda_s \to 0 \mathbf{c}$，它是下式的唯一解

$$\begin{bmatrix} q_{12} + \lambda & q_{21} \\ q_{12} & q_{21} + \lambda \end{bmatrix} \mathbf{c}_0 = \begin{bmatrix} q_{21}\pi_2 \\ q_{12}\pi_1 \end{bmatrix} + \lambda\, \mathbf{f}_{2,0}, \; \textit{where } \mathbf{f}_{2,0} \triangleq \lim_{\lambda_s \to 0} \mathbf{f}_2.$$

因此，极限 $\lambda_s \to 0\, c = \mathbf{1}^T\mathbf{c}_0$。

证明。在式 (35) 的线性方程组中令 μ 任意趋近于 0。极限 $\mathbf{f}_{2,0}$ 可通过在定义 $\mathbf{f}_2$, 的递推关系 (29) 中令 $\alpha_k = \frac{k\lambda_s}{n} = 0$ 得到。 □

直观而言，推论 2 可以解释为：当网络中没有新的数据包进入网络时，闲谈事件会主导网络中的信息流。在这种情况下，每个节点的准确度完全取决于分析子集中最新节点的准确度，因为如果没有新的数据包从源到达，该节点的内容必定会被所有其他节点接受。在这种情况下，尽管源没有向节点传输新鲜或准确的信息，但由于源处发生了状态翻转，节点仍可能持有准确信息。

现在我们刻画节点在没有速率限制的情况下可以相互交换信息时，平均准确度会如何变化。

推论 3。固定 $\lambda_s > 0$ （等价地，固定 $\mu > 0$）。令 $\mathbf{c}$ 为 (35)的解。那么，当 λ 任意增大时，有

$$\lim_{\lambda \to \infty} \mathbf{c} = \lim_{\lambda \to \infty} \mathbf{f}_2 = \begin{bmatrix} f_n^{(1)} \\ f_n^{(2)} \end{bmatrix}, \qquad \lim_{\lambda \to \infty} c = \lim_{\lambda \to \infty} \mathbf{1}^T\mathbf{c} = f_n,$$

其中， $f_n^{(1)}$和 $f_{(2)n}$ 由(34)给出，而 f_n 可由这两个量相加得到。

证明。 将 (35) 两边除以 λ，得到:

$$\left(I_2 + \frac{1}{\lambda}\begin{bmatrix} q_{12} + \mu & q_{21} \\ q_{12} & q_{21} + \mu \end{bmatrix}\right)\mathbf{c} = \mathbf{f}_2 + \frac{1}{\lambda}\mathbf{b}.$$

令 λ 趋于无穷大，得到 $\lim_{\lambda \to \infty}\mathbf{c} = \mathbf{f}_2$。此外，在 (29) 给出的 $\mathbf{f}_k$ 的递推关系中，对于每个 $k \in \{1, \ldots, n-1\}$，令 λ，从而 β_k，任意增大，会迫使 $\mathbf{f}_k - \mathbf{f}_{k+1}$ 趋于 0；反复迭代可得 $\lim_{\lambda \to \infty}\mathbf{f}_2 = \mathbf{f}_n$。最后，$\lim_{\lambda \to \infty} c = \mathbf{1}^T\mathbf{f}_n = f_n$。 □

随着 λ 增大，每个节点几乎可以瞬间与另一个节点通信。因此，推论 3 表明，整个网络的行为如同单个节点。由于流言协议基于新鲜度，平均准确度完全由网络中最新节点的准确度决定，该节点的信息会瞬间传播至整个网络。最后，我们在下面的推论中探讨节点完全停止信息扩散的情形。

推论 4. 固定 $\lambda_s > 0$ （等价地， $\mu > 0$）。在无信息传播极限下，当 λ趋近于0时，模态标记准确度向量满足

$$\begin{bmatrix} q_{12} + \mu & q_{21} \\ q_{12} & q_{21} + \mu \end{bmatrix} \lim_{\lambda \to 0} \mathbf{c} = \begin{bmatrix} q_{21}\pi_2 + \mu\pi_1 \\ q_{12}\pi_1 + \mu\pi_2 \end{bmatrix}, \quad (45)$$

并具有以下闭式解

$$\begin{aligned} \lim_{\lambda \to 0} \mathbf{c} &= \frac{1}{s(s+\mu)}\begin{bmatrix} q_{21}(q_{21} + \mu) \\ q_{12}(q_{12} + \mu) \end{bmatrix}, \\ \lim_{\lambda \to 0} c &= \lim_{\lambda \to 0} \mathbf{1}^T\mathbf{c} = \frac{\mu(q_{12} + q_{21}) + q_{12}^2 + q_{21}^2}{(q_{12} + q_{21})(q_{12} + q_{21} + \mu)}. \end{aligned} \quad (46)$$

证明。当 (35) 中的 λ 无限趋近于 0 时， W 中的 λI_2项消失，(35) 的右侧简化为$\mathbf{b}$。若使用 2×2 线性方程组，即可得到 $\pi_1 = \frac{q_{21}}{q_{12} + q_{21}}$, $\pi_2 = 1 - \pi_1$ 中的闭式表达式。 □

直观而言，如果网络中没有闲谈事件，每个节点的准确度将完全取决于源在不同状态之间跃迁的速度，以及源向各节点发送信息的速度。除了这两种

源动态之外，准确的信息，或者说任何类型的信息，都无法在网络中传播。

备注 1. 对于单个节点，也就是说，对于大小为$k = 1$的子集 A_k，最新节点的准确度等于平均准确度，因为该子集中只有一个节点。这意味着，在稳态条件下，$f_1 = c_1$。此外，正如前文所述，由于 $c_1 = c_k$，$\forall k$，求解 f_1 的递推关系即可完整刻画网络中的平均准确度。为求完整性，并给出推论 *1–4* 中明确的渐近极限，我们在本节中进行分析。

接下来，我们考虑一种特殊情形：二进制连续时间马尔可夫链是对称的，因此可以将本节推导出的矩阵值递推关系简化为标量递推关系。

C. 对称二进制源连续时间马尔可夫链的特殊情形

在本小节中，我们考虑源的连续时间马尔可夫链为对称情形，其生成矩阵为

$$\mathbf{Q} = \begin{bmatrix} -q & q \\ q & -q \end{bmatrix}.$$

利用该转移矩阵，很容易看出，对称连续时间马尔可夫链的稳态分布等于 $\pi_1 = \pi_2 = 0.5$。

对于这一特殊情况，可以将源视为按照速率为λ'_E 的泊松过程生成状态更新，该速率由跃迁速率得到，具体如下：

$$\lambda'_E = \pi_1\, q + \pi_2\, q = q. \tag{47}$$

因此，可以对 SHS 模型进行简化，使其符合以下形式（$\mathbf{C}(t)$, $\mathbf{X}(t)$) $\in \mathbb{R}^{2n}$，其中 $\mathbf{C}(t) = [C_1(t)\ .....,\ C_n(t)]$ 以及 $\mathbf{X}(t) = [X_1(t)\ .....,\ X_n(t)]$。换言之，当源的连续时间马尔可夫链具有对称性时，我们不需要像前几小节那样显式跟踪源的状态 $Q(t)$。这一新构造的系统遵循微分方程（$\mathbf{C}(t)$, $\mathbf{X}(t)) = \mathbf{0}_{2n}$，并且只有一个离散模式。转移集合简化为

$$\mathcal{E} = \{(0,0)\} \cup \{(0,j) : j \in \mathcal{N}\} \cup \{(i,j) : i,j \in \mathcal{N}\}, \tag{48}$$

其中包括：*i*) 表示源处状态翻转的事件 (0，0)，*ii*) 表示源向节点 j 发送更新的事件 (0，j)，以及 *iii*) 表示节点 i 向节点 j发送更新的事件 (i，j)，即节点 i 和 j之间的信息扩散。这些事件的速率为

$$\lambda_{ij} = \begin{cases} \lambda'_E, & i = 0,\ j = 0, \\ \dfrac{\lambda_s}{n}, & i = 0,\ j \in \mathcal{N}, \\ \dfrac{\lambda}{n-1}, & i,j \in \mathcal{N}. \end{cases} \tag{49}$$

现在，我们可以分析对称二元信息源下集合A_k 中最新节点的准确度，记为 $F_{A_k,\ \mathrm{sym}}$；以及集合 A_k 的准确度，记为$C_{A_k,\ \mathrm{sym}}$。由于信息源状态转移速率对称，我们无需考虑它们对应的模式标记形式。

利用式 (16) 和 (19) 中给出的重置映射，以及随机混合系统生成元恒等式，可得

$$\sum_{(i,j)\in\mathcal{E}} \lambda_{ij}\Big(\mathbb{E}[\psi(\phi_{i,j})] - \mathbb{E}[\psi]\Big) = 0. \tag{50}$$

然后，我们可以得到稳态值 $f_{k,\mathrm{sym}} = \lim_{t\to\infty} \mathbb{E}[F_{A_k,\mathrm{sym}}(t)]$ 如下：

$$\lambda'_E(1-2f_{k,\mathrm{sym}}) + \alpha_k(1-f_{k,\mathrm{sym}}) + \beta_k(f_{k+1,\mathrm{sym}} - f_{k,\mathrm{sym}}) = 0, \tag{51}$$

其中，α_k 和 β_k 在（25）中定义。最后，重新整理（51）中的各项，可得到如下向后递归关系：

$$f_{k,\mathrm{sym}} = \frac{\lambda'_E + \alpha_k + \beta_k f_{k+1,\mathrm{sym}}}{2\lambda'_E + \alpha_k + \beta_k}. \tag{52}$$

接下来，定义 $C_{\emptyset,\mathrm{sym}} = 0$，由于空集不可能包含准确信息，经过类似的推导可得到 $C_{A_k,\mathrm{sym}}$ 的稳态刻画，即 $c_{k,\mathrm{sym}} = \lim_{t\to\infty} \mathbb{E}[C_{A_k,\mathrm{sym}}(t)]$ 此外，我们将 $F_{\{i,j\},\mathrm{sym}}$ 的稳态值记为 $f_{2,\mathrm{sym}} = \lim_{t\to\infty} \mathbb{E}[F_{\{i,j\},\mathrm{sym}}(t)]$。因此，我们有：

$$\lambda'_E(1-2c_k) + \lambda k\left(\frac{f_{2,\mathrm{sym}} - c_{1,\mathrm{sym}}}{k}\right) + \alpha_k\left(\frac{1-c_{1,\mathrm{sym}}}{k}\right) = 0.$$

汇总 $c_{k,\mathrm{sym}}$ 中的所有项，并求解 $c_{k,\mathrm{sym}}$ 以及$f_{1,\mathrm{sym}} = c_{1,\mathrm{sym}} = c_{k,\mathrm{sym}}$，可得

$$c_{k,\mathrm{sym}} = \frac{\lambda'_E + \frac{\lambda_s}{n} + \lambda f_{2,\mathrm{sym}}}{2\lambda'_E + \frac{\lambda_s}{n} + \lambda}, \tag{53}$$

对于所有 $k \geq 1$。由（52）和（53）可以看出，基于新鲜度的准确度和平均准确度的递推公式都化为标量值公式，而非矩阵值公式，因为无需通过指示函数显式跟踪源状态。接下来，我们将针对具有一般 $M > 2$ 状态连续时间马尔可夫链的源，扩展准确度分析。

IV. 一般多状态信息源

现在考虑信息源具有一个包含 $M > 2$ 个状态的连续时间马尔可夫链的情形，即 $Q(t) \in \{1, \cdots, M\}$。与二元情形不同，在信息源反转后，我们不再能假设平均准确度 C_{A_k} 和基于新鲜度的准确度 F_{A_k}都会简单地翻转，因为信息源不再反转，而是转变为另一个状态。因此，在信息源发生跃迁后，集合 A_k 中最新节点的准确度，即 F_{A_k}，取决于最新节点所持有的内容是否与信息源跃迁至的下一个状态一致，即$Q'(t)$。因此，根据新的信息源定义，$F_{A_k}(t)$ 可以改写如下：

$$F_{A_k}(t) \triangleq \mathbb{1}_{\{R_k(t)=Q(t)\}}, \tag{54}$$

其中，$R_k(t) = S_{\arg\text{最小值}_{j\in A_k} X_j(t)}$ 表示网络中由 k 个节点组成的子集 A_k 内最新节点所持有的内容。接下来，我们定义一个新的联合过程：

$$Z_k(t) \triangleq \big(Q(t),\, R_k(t)\big), \tag{55}$$

其取值范围为有限状态空间

$$\mathcal{S} \triangleq \{(q,r) : q \in \{1,\ldots,M\},\ r \in \{1,\ldots,M\}\}. \tag{56}$$

令 $\pi_q^{(k)},\ r$ 表示 $Z_k(t)$ 的平稳分布，该分布为

$$\pi_{q,r}^{(k)} \triangleq \mathbb{P}(Q=q,\ R_k=r), \tag{57}$$

对于所有（$q,\ r$）$\in \mathcal{S}$，所得的稳态新鲜度准确率 $f_k = \lim_{t\to\infty}\mathbb{E}[F_{A_k}(t)]$，可通过以下方式获得：

$$f_k \triangleq \mathbb{P}(R_k=Q) = \sum_{q=1}^{M}\pi_{q,q}^{(k)}. \tag{58}$$

对于每个 $k \in \{1,\dots,n\}$，过程 $Z_k(t)$ 是定义在 $\mathcal{S}$ 上的连续时间马尔可夫链。我们将 Z_k（t）的状态列为 {（1, 1）、（1, 2）⋯、（1, M）、（2, 1）⋯、（2, M）⋯、（M, M）}，并定义一个生成矩阵 Q_k，其维度为 $M^2 \times M^2$。Q_k中的转移源于三种情况：i）$Q(t)$的源到源转移，ii）源到A_k的更新事件，以及 iii）源自 A_k之外节点的信息传播事件。因此，矩阵 Q_k 将由三个不同的组成部分构成，即 Q_k^{src}、Q_k^{push} 和 Q_k^{out}，它们的维度分别为 $M^2 \times M^2$。下面我们描述每种情况的贡献。

从状态 i 到状态 j 且带有 $i \neq j$ 的信息源状态转移不会改变 $R_k(t)$。因此，生成矩阵的源分量Q_k^{src}中相应的对角元素和非对角元素由下式给出

$$Q_k^{\text{src}}\big((i,r),(j,r)\big) = \begin{cases} q_{ij}, & \text{for } j\neq i,\\ -\sum_{\substack{j=1\\ j\neq i}}^{M} q_{ij}, & \text{for } j=i, \end{cases} \tag{59}$$

对所有 r，Q_k^{src}的其余元素均等于0。我们还注意到，Q_k^{src}的每一行之和为零。

接下来，从源向集合 A_k 推送的总更新率为 α_k，如(25)所述。一次推送事件将最新内容设置为等于源状态。如果最新内容已经等于源状态，则不会发生变化。因此，生成矩阵的推送分量Q_k^{push} 具有非对角元素

$$Q_k^{\text{push}}\big((i,r),(i,i)\big) = \alpha_k\,\mathbb{1}_{\{r\neq i\}}, \tag{60}$$

以及对角元素

$$Q_k^{\text{push}}\big((i,r),(i,r)\big) = -\alpha_k\,\mathbb{1}_{\{r\neq i\}}, \tag{61}$$

并且 Q_k^{push} 的其余元素给定为0。从 A_k 外部到达该子集的闲谈事件以速率β_k发生。与二元情形一样，我们需要扩大节点集合，以便在信息传播事件下获得 F_{A_k} 的刻画。当我们将所关注的集合从 k扩大到 $k+1$ 个节点时，最新节点的内容根据给定源状态下 R_{k+1} 的条件分布从 r 变为 r'。因此，我们定义条件概率

$$p_{r|q}^{(k+1)} \triangleq \mathbb{P}(R_{k+1}=r \mid Q=q), \tag{62}$$

对于 $q,r\in\{1,\dots,M\}$。利用（$k+1$）节点过程的平稳分布，我们有

$$p_{r|q}^{(k+1)} = \frac{\pi_{q,r}^{(k+1)}}{\pi_q^{(k+1)}}, \tag{63}$$

算法 1 反向构造 $\{Q_k,\boldsymbol{\pi}^{(k)}\}_{k=}^{n}$ $_1$ and 计算 $\{f_k\}_{k=1}^{n}$

要求： 网络规模 n；源连续时间马尔可夫链状态数 M，源推送速率 α_k，Gossip 速率 β_k；稳态方程 (67)；条件定义 (62)–(63)。

确保： 生成矩阵 $\{Q_k\}_{k=1}^{n}$，平稳分布 $\{\boldsymbol{\pi}^{(k)}\}_{k=1}^{n}$，以及准确率 $\{f_k\}_{k=1}^{n}$。

1: **在** $k=n$ **初始化：** 设置 $\beta_n \leftarrow 0$。
2: 构造 $Q_n \leftarrow Q_n^{\text{src}} + Q_n^{\text{push}}$。
3: 对 $k=n$ 求解 (67)，得到 $\boldsymbol{\pi}^{(n)}$。
4: **对于** $k=n-1,n-2,\dots,1$ **do**
5: 使用 $\boldsymbol{\pi}^{(k+1)}$，计算条件概率 $p_{r|q}^{(k+1)}$ 通过 (62)–(63)。
6: 利用 $p_{r|q}^{(k+1)}$ 构造 Q_k（以及模型构造 Q_k).
7: 求解 (67)，得到平稳分布 $\boldsymbol{\pi}^{(k)}$。
8: **结束**
9: **对于** $k=1$、2、.....、n **执行**
10: 利用(58)由 $\boldsymbol{\pi}^{(k)}$ 计算 f_k。
11: **结束循环**
12: **返回** $\{Q_k,\boldsymbol{\pi}^{(k)},f_k\}_{k=1}^{n}$。

其中 $\pi_q^{(k+1)} \triangleq \sum_{r'=1}^{M}\pi_{q\,r'}^{(k+1)}$.，然后，进入 A_k 的外部八卦会在生成矩阵的八卦分量 Q_k^{out} 中引发如下形式的转移

$$Q_k^{\text{out}}\big((q,r),(q,r')\big) = \beta_k p_{r'|q}^{(k+1)}, \qquad r'\neq r, \tag{64}$$

以及相应的对角元素

$$Q_k^{\text{out}}\big((q,r),(q,r)\big) = -\beta_k\sum_{\substack{r'=1\\ r'\neq r}}^{M} p_{r'|q}^{(k+1)} = -\beta_k\big(1-p_{r|q}^{(k+1)}\big). \tag{65}$$

综合这三种情形，Z的总生成矩阵 Q_k 由下式给出 $_k(t)$

$$Q_k = Q_k^{\text{src}} + Q_k^{\text{push}} + Q_k^{\text{out}}. \tag{66}$$

对于每个固定的 k，平稳分布 $\pi^{(k)}$ 满足

$$\boldsymbol{\pi}^{(k)}Q_k = \mathbf{0}, \tag{67}$$

以及 $\sum_{(q,r)\in\mathcal{S}}\pi_{q\ ,r}^{(k)} = 1$其中 $\boldsymbol{\pi}^{(k)}$ 是由 $\boldsymbol{\pi}^{(k)} = [\pi_{1,1}^{(k)},\cdots,\pi_{1,M}^{(k)},\pi_{2,1}^{(k)},\cdots,\pi_{2,M}^{(k)},\cdots,\pi_{M,1}^{(k)},\cdots,$ 组成 $\pi_{M,M}^{(k)}]$. $p_{r|q}^{(k+1)}$ 62 63 的行向量。由于（ ）–（ ）中的值取决于（$k+1$）节点系统的平稳分布，因此可以按照算法 1 递归地构建生成矩阵 $\{Q_k\}$ 及其对应的平稳分布。

对于每个 k，在获得平稳分布 $\pi_q^{(k)},_r$ 后，基于新鲜度的准确度 f_k 由式（58）给出。注意，与二元情形一样，任何大小为 k 的子集的平均准确度都与单个节点的准确度一致，即

$$f_1 = c_1 = c_k, \qquad \forall k\in\{1,\dots,n\}, \tag{68}$$

因此，该联合连续时间马尔可夫链模型刻画了网络中基数为 k的任意子集的平均准确度。

接下来，我们给出一个引理，说明网络的信息内容分布仅取决于源节点连续时间马尔可夫链的平稳分布。

引理 1. 在稳态条件下，对于任意给定的 λ、$\lambda_s > 0$，网络中任意基数为 k、持有信息 q 的子集中的节点期望数量等于 $k\pi_q$。

证明。考虑对式 (9)、(10) 和 (12) 中所述的 SHS 模型进行扩展，其中包含与 M 源节点状态相对应的 M 离散模式。

令 $\bar{\psi}_k(\mathbf{C},\mathbf{X}) = N_{A_k}^{(q)}$ 为时不变测试函数，用于给出集合 A_k 中持有信息 q 的节点数量，其形式化定义如下：

$$\bar{\psi}_k(\mathbf{C},\mathbf{X}) \triangleq N_{A_k}^{(q)}(t) = \sum_{j\in A_k} \mathbb{1}_{\{S_j(t)=q\}}, \tag{69}$$

那么，$N_{A_k}^{q}$ 的跃迁映射可写为：

$$N_{A_k}^{(q)\prime} = \begin{cases} N_{A_k}^{(q)} + \mathbb{1}_{\{Q=q\}} - H_{\{j\}}^{(q)}, & e=(0,j), j\in A_k, \\ N_{A_k}^{(q)} + H_{\{i,j\}}^{(q)} - H_{\{j\}}^{(q)}, & e=(i,j), i\in\mathcal{N}, j\in A_k, \\ N_{A_k}^{(q)}, & \text{otherwise}, \end{cases} \tag{70}$$

其中 $H_{A_k}^{(q)}(t) = \mathbb{1}_{\{S_{\arg\min_{j\in A_k} X_j(t)}=q\}}$。注意，对于仅由单个节点组成的子集，$H_{\{j\}A_k}(t)$ 可简化为 $H_j^{(q)}(t) = \mathbb{1}_{\{S_j(t)=q\}}$。这使我们能够使用 H_{A_k} 的递归表达式来建模式 (69) 中指示函数项的变化。为了刻画 $n_k^{(q)} = \lim_{t\to\infty}\mathbb{E}[N_{A_k}^{(q)}(t)]$，我们首先需要刻画 $h_k^{(q)} = \lim_{t\to\infty}\mathbb{E}[H_{A_k}^{(q)}(t)]$，这可以通过以下重置映射实现：

$$H_{A_k}^{(q)\prime} = \begin{cases} \mathbb{1}_{\{Q=q\}}, & e=(0,j),\ j\in A_k, \\ H_{A_{k+1}}^{(q)}, & e=(i,j),\ i\in\mathcal{N}\setminus A_k,\ j\in A_k, \\ H_{A_k}^{(q)}, & \text{otherwise}. \end{cases} \tag{71}$$

注意到，对于 $h_k^{(q)} = \lim_{t\to\infty}\mathbb{E}[\mathbb{1}_{\{Q=q\}}] = \pi_q$ 的稳态平衡方程可写为：

$$h_k^{(q)} = \frac{\dfrac{k\lambda_s}{n}\pi_q + \dfrac{k(n-k)\lambda}{n-1}h_{k+1}^{(q)}}{\dfrac{k\lambda_s}{n} + \dfrac{k(n-k)\lambda}{n-1}}. \tag{72}$$

利用递推关系在 $k=n$ 时化简为 $h_n = \pi_q$ 这一事实，可以证明，对于所有 $k\in\{1,\cdots,n\}$，均有 $h_k^{(q)} = \pi_q$。利用由 (72) 得出的结果，$n_k^{(q)}$ 的递推表达式可写为：

$$0 = \frac{k\lambda_s}{n}\big(\pi_q - h_1^{(q)}\big) + k\lambda\big(h_2^{(q)} - h_1^{(q)}\big). \tag{73}$$

根据 (69) 中 $N_{A_k}^{(q)}$ 的定义，我们有

$$n_k^{(q)} = \lim_{t\to\infty}\mathbb{E}[N_{A_k}^{(q)}] = \sum_{j\in A_k}\lim_{t\to\infty}\mathbb{E}[\mathbb{1}_{\{S_j=q\}}] = k\,h_1^{(q)}. \tag{74}$$

将此结果与 $h_1^{(q)} = \frac{n_k^{(q)}}{k}$ 一并代入 (73)，得到

$$0 = \frac{k\lambda_s}{n}\left(\pi_q - \frac{n_k^{(q)}}{k}\right) + k\lambda\big(h_2^{(q)} - h_1^{(q)}\big), \tag{75}$$

$$n_k^{(q)} = k\pi_q + \frac{nk\lambda}{\lambda_s}\big(h_2^{(q)} - h_1^{(q)}\big). \tag{76}$$

最后，将 $h_2^{(q)} = h_1^{(q)} = \pi_q$ 代入 (76)，得到

$$n_k^{(q)} = k\pi_q. \tag{77}$$

至此，证明完毕。 □

V. 新鲜度导致的信息准确性

如前所述，一个节点可以通过两种不同方式拥有准确信息。首先，节点可能持有源处生成的最新更新，此时其版本年龄为零。其次，即使节点没有最新鲜的更新，即其版本年龄严格为正，它仍可能是准确的，因为源处随后的信息翻转会使源的当前状态与节点所存储的信息一致。在这种情况下，尽管就版本年龄而言已经过时，节点的信息仍然准确。在本节中，我们区分这两种情形，并量化由此产生的、可归因于*i*)新鲜度效应和 *ii*)源状态转移的信息不准确度。

定理 2. 网络中任意一个基数为 $k\in\{1,2,\cdots,n-1\}$ 的子集内，持有严格新鲜且准确的信息的节点比例为 $G_k = \sum_{m=1}^{M} c_k^{(m)} g_k^{(m)}$，其中 $M\in\{2,3,\ldots\}$，且 $g_k^{(m)} = \frac{\alpha_k\pi_m + \beta_k g_{k+1}^{(m)}}{\nu_m + \alpha_k + \beta_k}$。

证明。首先，我们采用与第三节类似的方式，使用 SHS 方法研究二元源连续时间马尔可夫链中由新鲜度和源状态翻转带来的平均准确度。随后，我们针对一般多状态信息源，用 M 个不同的测试函数扩展这一 SHS 模型。对于二元情形，考虑两个新的带模式标签的测试函数，明确建模有效的信息新鲜度（或类似地，源状态翻转）对准确信息传播的作用。我们定义 $\tilde{\psi}_k^{(m)}(Q,\mathbf{Z}) = c_k^{(m)}\mathbb{1}_{\{Q=m,X_{A_k}=0\}}$，表示集合 A_k 因新鲜信息而具有的平均准确度，其稳态值记为：

$$\lim_{t\to\infty}\mathbb{E}\big[\tilde{\psi}_k^{(m)}\big] = c_k^{(m)} g_k^{(m)}, \tag{78}$$

其中 $g_k^{(m)} \triangleq \lim_{t\to\infty}\mathbb{E}\left[\mathbb{1}_{\{Q=m,X_{A_k}=0\}}\right] = \mathbb{P}(Q=m,\ X_{A_k}=0)$ 由于 $c_k^{(m)}$ 是常数（已由信息准确度分析确定），因此只需刻画 $g_k^{(m)}$。由于对这两个测试函数的分析相似，这里我们只提供第一个带模式标记函数的分析，方式与带模式标记的新鲜度函数分析类似。

考虑跃迁 $e=(0,0,1\to 2)$。跃迁前，$\tilde{\psi}_k^{(1)}(Q,\mathbf{Z}) = \mathbb{1}_{\{Q=1,X_{A_k}=0\}}$ 的求值结果为 $\tilde{\psi}_k^{(1)}(Q,\mathbf{Z}) = \mathbb{1}_{\{X_{A_k}=0\}}$，因为源内容指示变量取值为 1。上述跃迁后，网络中所有节点的版本年龄都增加 1，如式（15）中的跃迁映射所示。因此，版本年龄的指示函数会使该表达式的求值结果变为零，从而产生由 $-\mathbb{1}_{\{Q=1,X_{A_k}=0\}}$ 项给出的总体变化。

现在，考虑由 $(0,j)$ 表示的源推送跃迁，适用于 $j\in A_k$。在此类跃迁之后，接收节点将包含准确且新鲜的信息，因此测试函数等于 $\tilde{\psi}_k^{(1)}(Q,\mathbf{Z}) = \mathbb{1}_{\{Q=1\}}$，从而使该跃迁下的总体变化为 $\mathbb{1}_{\{Q=1\}} - \mathbb{1}_{\{Q=1,X_{A_k}=0\}}$。

最后，考虑由跃迁 $e=(i,j)$ 给出的信息传播事件，其中 $j\in A_k$ 和 $i\in\mathcal{N}\setminus A_k$。在此跃迁下，源内容不变，版本年龄分析必须扩展到超集 A_{k+1}，如(15)所示。

因此，测试函数在此跃迁下的总体变化将为
$\mathbb{1}_{\{Q=1\}}\big(\mathbb{1}_{\{X_{A_{k+1}}=0\}}-\mathbb{1}_{\{X_{A_k}=0\}}\big)$。

对所有这些项取期望，并分别加上速率为 α_k 和 β_k 的源推送事件和外部八卦事件，可得以下 SHS 平衡方程：

$$g_k^{(1)}=\frac{\alpha_k\pi_1+\beta_k g_{k+1}^{(1)}}{q_{12}+\alpha_k+\beta_k}. \tag{79}$$

通过类似的推导，可得第二模式对应项的表达式如下：

$$g_k^{(2)}=\frac{\alpha_k\pi_2+\beta_k g_{k+1}^{(2)}}{q_{21}+\alpha_k+\beta_k}. \tag{80}$$

定义 $\mathbf{g}_k\triangleq[g_k^{(1)},g_k^{(2)}]^T$ 和 $\boldsymbol{\pi}\triangleq[\pi_1,\pi_2]^T$. 于是，这两个方程可以像之前的测试函数一样，写成矩阵递推的形式：

$$\mathbf{g}_k=\mathbf{A}_k\,\boldsymbol{\pi}+\mathbf{B}_k\,\mathbf{g}_{k+1}, \tag{81}$$

其中

$$\mathbf{A}_k\triangleq\begin{bmatrix}\frac{\alpha_k}{q_{12}+\alpha_k+\beta_k} & 0\\ 0 & \frac{\alpha_k}{q_{21}+\alpha_k+\beta_k}\end{bmatrix},\quad \mathbf{B}_k\triangleq\begin{bmatrix}\frac{\beta_k}{q_{12}+\alpha_k+\beta_k} & 0\\ 0 & \frac{\beta_k}{q_{21}+\alpha_k+\beta_k}\end{bmatrix}.$$

至此，二元源情形的分析告一段落。该分析可推广到具有 M 个状态的一般信息源，其中包含 M 个不同的带模式标签的测试函数，形式为 $\tilde{\psi}_k^{(m)}(Q,\mathbf{X})=c_k^{(m)}\mathbb{1}_{\{Q=m,X_{A_k}=0\}}$，$m\in\{1,\dots,M\}$ 对于。

根据式 (79) 和 (80) 中递归表达式所呈现的规律，可以使用这 M 个不同的测试函数将推导推广如下：

$$g_k^{(m)}=\frac{\alpha_k\pi_m+\beta_k g_{k+1}^{(m)}}{\nu_m+\alpha_k+\beta_k},\qquad k=1,\dots,n-1. \tag{82}$$

因此，任意 k 个节点子集中持有准确且新鲜信息的节点比例 $G_k=\sum_{m=1}^{M}c_k^{(m)}g_k^{(m)}$，$M\in\{2,3,\dots\}$，得到，由此完成证明。

□

现在我们陈述几个由定理 2 直接推导出的极限，首先从网络中准确但陈旧节点的比例开始。

推论 5。任意 k 个节点子集中拥有准确但陈旧信息的节点比例为 c_k-G_k，其中 c_k 的值可通过遵循第三节和第四节中的分析得到。

证明。回顾可知，c_k 表示集合 A_k 持有准确信息的总体概率（比例），即 $c_k=\mathbb{P}(R_k=Q)$，而 G_k 表示其同时准确且严格新鲜的概率，即 $G_k=\mathbb{P}(R_k=Q,\ X_{A_k}=0)$。由于事件 $\{X_{A_k}=0\}$ (新鲜) 和 $\{X_{A_k}>0\}$ (陈旧) 互不相交且完备，因此有
$\{R_k=Q\}=\{R_k=Q,\,X_{A_k}=0\}\cup\{R_k=Q,\,X_{A_k}>0\}$.取概率可得 $c_k=G_k+\mathbb{P}(R_k=Q,\,X_{A_k}>0)$，因此，准确但陈旧的节点所占比例为 $\mathbb{P}(R_k=Q,\,X_{A_k}>0)=c_k-G_k$.

□

现在，我们来探究如果源可以不受速率限制地向网络推送更新，新鲜度对信息准确性的影响。

推论 6. 对于任意 $k\in\{1,\dots,n\}$ 和模式 $m\in\{1,\dots,M\}$，如果 λ_s（因此 α_k）无限增大，那么

$$\lim_{\lambda_s\to\infty}g_k^{(m)}=\pi_m,\ \lim_{\lambda_s\to\infty}G_k=\sum_{m=1}^{M}c_k^{(m)}\pi_m=\sum_{m=1}^{M}\pi_m^2. \tag{83}$$

证明。由 (82) 可得

$$g_k^{(m)}=\frac{\alpha_k\pi_m+\beta_k g_{k+1}^{(m)}}{\nu_m+\alpha_k+\beta_k}.$$

将分子和分母除以 α_k，并令 α_k 趋于无穷大，取其极限：

$$\lim_{\alpha_k\to\infty}g_k^{(m)}=\lim_{\alpha_k\to\infty}\frac{\pi_m+(\beta_k/\alpha_k)g_{k+1}^{(m)}}{(\nu_m/\alpha_k)+1+(\beta_k/\alpha_k)}=\pi_m,$$

因为 $0\le g_{k+1}^{(m)}\le 1$。关于 $G_k=\sum_m c_k^{(m)}g_k^{(m)}$ 的结论可由 (44) 中给出的 c_k 在 λ_s 趋于无穷大时的极限得出。

□

直观地说，当源以任意快的速度推送时，节点几乎必然是新鲜的，因此事件 $\{X_{A_k}=0\}$ 不再具有筛选作用，而模式概率由源的稳态行为主导 π_m。

现在我们来研究源何时停止更新网络。

推论 7. 对于任意 $k\in\{1,\dots,n\}$ 和 $m\in\{1,\dots,M\}$，有

$$\lim_{\lambda_s\to0^+}g_k^{(m)}=0,\qquad \lim_{\lambda_s\to0^+}G_k=0. \tag{84}$$

证明。在 $k=n$ 处没有外部节点，因此 $\beta_n=0$，递推关系简化为

$$\lim_{\alpha_n\to0}g_n^{(m)}=\lim_{\alpha_n\to0}\frac{\alpha_n\pi_m}{\nu_m+\alpha_n}=0.$$

现在对 $k=n-1,n-2,\dots,1$ 反向应用 (82)。代入 $\alpha_k=0$ 后，递推关系变为

$$g_k^{(m)}=\frac{\beta_k}{\nu_m+\beta_k}\,g_{k+1}^{(m)},$$

且乘法因子 $\beta_k/(\nu_m+\beta_k)$ 有限。由于 $g_n^{(m)}$ 趋近于 0，因此所有先前的 $g_k^{(m)}$ 也趋近于 0，故 G_k 趋近于 0。

□

直观而言，随着网络与源断开，所有节点的版本年龄都会趋于无穷大，因此零年龄指示变量 $\mathbb{1}_{\{X_{A_k}=0\}}$ 会使 $g_k^{(m)}=\mathbb{P}(Q=m,X_{A_k}=0)$ 趋于 0。

推论 8 描述了当所有节点都瞬时进行信息传播时 G_k 的行为，即当 $\lambda\to\infty$ 趋于极限时的行为。

推论 8. 对于任意 $\lambda_s>0$（因此，$\alpha_k>0$）以及任意 $m\in\{1,\cdots,M\}$，

$$\lim_{\lambda\to\infty}\big(g_k^{(m)}-g_{k+1}^{(m)}\big)=0,\qquad k=1,\dots,n-1, \tag{85}$$

因此，当 λ 变得很大时，$g_1^{(m)}=\cdots=g_n^{(m)}$ 趋于极限。

证明. 由 (82) 可得 $g_k^{(m)}=\frac{\alpha_k\pi_m+\beta_k g_{k+1}^{(m)}}{\nu_m+\alpha_k+\beta_k}$. 接下来，我们定义 $D_k^{(m)}\triangleq\nu_m+\alpha_k+\beta_k$，$w_k^{(m)}\triangleq\frac{\alpha_k}{D_k^{(m)}}$，$v_k^{(m)}\triangleq\frac{\beta_k}{D_k^{(m)}}$，以及 $r_k^{(m)}\triangleq\frac{\nu_m}{D_k^{(m)}}$. 然后，我们可以将 $g_k^{(m)}$ 写作

$$g_k^{(m)}=w_k^{(m)}\pi_m+v_k^{(m)}g_{k+1}^{(m)}+r_k^{(m)}\cdot 0,$$

使用 $w_k^{(m)}+v_k^{(m)}+r_k^{(m)}=1$，我们得到

$$g_k^{(m)}-g_{k+1}^{(m)}=w_k^{(m)}\Big(\pi_m-g_{k+1}^{(m)}\Big)-r_k^{(m)}g_{k+1}^{(m)}.$$

使用 $0 \le g_{k+1}^{(m)} \le 1$ 和 $0 \le \pi_m \le 1$，我们得到

$$\left| g_k^{(m)} - g_{k+1}^{(m)} \right| \le w_k^{(m)} + r_k^{(m)} = \frac{\alpha_k + \nu_m}{\nu_m + \alpha_k + \beta_k}.$$

随着 λ（因此 β_k）任意增大，右端趋于 0，因此我们有 $\lim_{\lambda\to\infty}\left(g_k^{(m)} - g_{k+1}^{(m)}\right) = 0$. □

直观而言，当信息扩散速度任意快时，子集 A_k 与其超集之间的区别便会消失，因此从新鲜度的角度来看，网络的行为就像单个节点，这与前文各节的发现一致。

最后，我们考察网络内的信息扩散停止、仅剩源推送的情形。

推论 9. 固定 $\lambda_s > 0$，并考虑 $\lambda = 0$（因此，对于所有 k，$\beta_k = 0$）。那么，对于每个 $k \in \{1,\dots,n\}$ 和 $m \in \{1,\dots,M\}$，我们有

$$\lim_{\lambda\to 0} g_k^{(m)} = \mathbb{P}(Q = m,\ X_{A_k} = 0) = \frac{\alpha_k\,\pi_m}{\nu_m + \alpha_k}. \tag{86}$$

因此，

$$\lim_{\lambda\to 0} G_k = \sum_{m=1}^{M} \lim_{\lambda\to 0} c_k^{(m)}\, g_k^{(m)} = \sum_{m=1}^{M} \frac{\alpha_k\,\pi_m}{\nu_m + \alpha_k} \lim_{\lambda\to 0} c_k^{(m)}. \tag{87}$$

证明。当 $\lambda = 0$ 时，不会发生信息扩散转移。随着其任意趋近于 0，(82) 得出

$$\lim_{\lambda_k\to 0} g_k^{(m)} = \frac{\alpha_k \pi_m}{\nu_m + \alpha_k}, \qquad k = 1,\dots,n,$$

由此得证。将此表达式代入 (86) 即得 $\lim_{\lambda\to 0} G_k = \sum_{m=1}^{M} c_k^{(m)} g_k^{(m)}$（）。由于一般的 M 状态信息源不存在 $c_k^{(m)}$ 的闭式解，极限 $\lim_{\lambda\to 0} c_k^{(m)}$ 可通过在算法 1 中设置任意低的 Gossip 速率 λ 并求解递推关系来计算。若源节点连续时间马尔可夫链是二元的，则 $\lim_{\lambda\to 0} c_k^{(m)}$ 存在闭式解，并由推论 4 给出。 □

当 $\lambda = 0$ 时，A_k 中的节点只能通过直接源推送变为新鲜，因为不存在点对点传播来从该集合外部“引入”新鲜度。因此，$g_k^{(m)} = \mathbb{P}(Q = m, X_{A_k} = 0)$ 等于源处于模式 m（即 π_m）的时间稳态占比，乘以在源离开模式 m 之前推送到达的概率，即 $\alpha_k/(\nu_m + \alpha_k)$。

在下一节中，我们将给出数值结果，并将第 III–V 节推导出的理论结果与仿真进行比较。

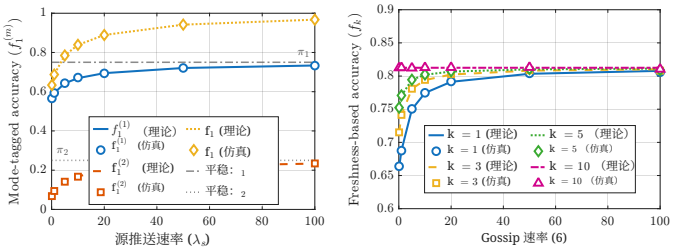
VI. 数值结果与渐近分析见解

本节将给出以下方面的数值结果：a）二元信息源；b）多状态信息源；以及 c）新鲜度和信息源反转对信息准确性的影响比较。

A. 二元信息源的结果

在我们的第一组仿真结果中，我们对第二节所述的信息传播网络进行了仿真，共运行 2，250，000 个时间步，其中包含 $n = 10$ 个节点和 $M = 2$ 个源状态，并分别遍历 λ 和 λ_s 的取值，同时保持其他变量不变。源生成器矩阵选取为

$$\mathbf{Q} = \begin{bmatrix} -0.25 & 0.25 \\ 0.75 & -0.75 \end{bmatrix}.$$



(a) (b)

图2. 基于新鲜度的准确度与(a) λ_s（针对 $k \in \{1,3,5,10\}$）和(b) λ（针对 $k = 1$）的关系。

在图 2(a) 中，我们展示了节点的新鲜度准确度（f_1）以及基于新鲜度的模式标记准确度，即 $f_1^{(m)}$，随 λ_s 的变化情况。在图 2(b) 中，我们展示了基于新鲜度的准确度，即 f_k，对于 $k \in \{1,3,5,10\}$ 随 λ 的变化情况。我们注意到，对于任意基数为 k 的子集，f_1 还给出了平均准确率 $c_k = c_1$。在每幅图中，未被遍历的参数均固定为 1；也就是说，图 2(a) 中为 $\lambda = 1$，图 2(b) 中为 $\lambda_s = 1$。为了理解图 2(a)，我们将注意力集中在式 (29) 中给出的递归 $\mathbf{f}_k$ 推导，以及式 (30) 中的 W_k 和 $\mathbf{v}_k$ 表达式上。随着 λ_s 增大，α_k 也按比例增加，这使得矩阵 W_k 趋近于 $\alpha_k I_2$，而向量 $\mathbf{v}_k$ 越来越接近 $\alpha_k\boldsymbol{\pi}$。由于式 (29) 的最后一项，即 $V_k\mathbf{f}_{k+1}$（其中 $V_k = \beta_k I_2$），仅由有限元素组成，因此随着 λ_s 增大，向量 $\mathbf{f}_k$ 收敛到 $\boldsymbol{\pi}$。直观地说，如果源始终将其当前信息推送到网络中，节点就始终携带新鲜且准确的信息，因此 $f_k^{(1)}$ 和 $f(2)_k$ 将分别对其模式标签取源的平稳分布。此外，由于 $f_k = f(1_k + f_k^{(2)})$，在极限情况下 $\lim_{\alpha_k\to\infty} f_k = \pi_1 + \pi_2 = 1$，这意味着任意子集中的最新节点始终是准确的。这一趋势在图 2(a) 中得到了观察，并由推论 1 予以证明。随着 λ 增大，从而 β_k 按比例增大，我们有 $\frac{1}{\beta_k} W_k$ 收敛到 I_2，且 $\frac{1}{\beta_k}\mathbf{v}_k$ 越来越接近 0。因此，我们有 $\lim_{\beta_k\to\infty} \mathbf{f}_k = \mathbf{f}_{k+1}$。直观地说，如果所有节点在每个时间瞬间都相互通信，则整个网络的行为就如同一个单一节点。这一趋势可见于图 2(b)，并由推论 3 予以证明。

从图2(a)和图2(b)可以看出，增大 λ_s对提高平均准确度的效果远大于增大 λ。其原因在于，当 λ 增大而 λ_s 保持不变时，节点能够获得网络中最新鲜的信息，但网络中最新鲜的信息将取决于源的速率 λ_s。另一方面，如果增大 λ_s 而保持 λ 不变，网络中的每个节点仍然能够从源获得最新的信息；因此，节点的准确度收敛到1。

我们的第二组仿真结果如图3(a)和图3(b)所示，说明了对称二元信源下平均准确度的行为。与先前的仿真结果类似，未被扫描的参数固定为1，即图3(a)中为$\lambda = 1$，图3(b)中为 $\lambda_s = 1$。这些图验证了如下稳态假设：所有

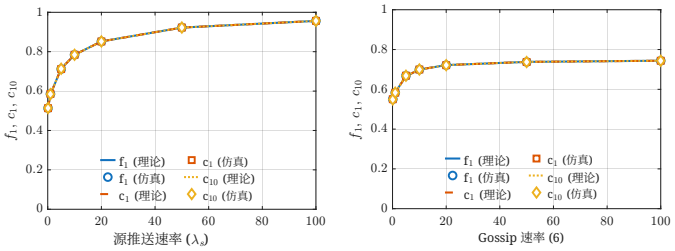


图3. 对称连续时间马尔可夫链的节点基于新鲜度的准确度和平均准确度（分别为 $f_{1,\ \mathrm{sym}}$ 和 $c_{1,\ \mathrm{sym}}$）与 λ_s 和 λ 的关系。

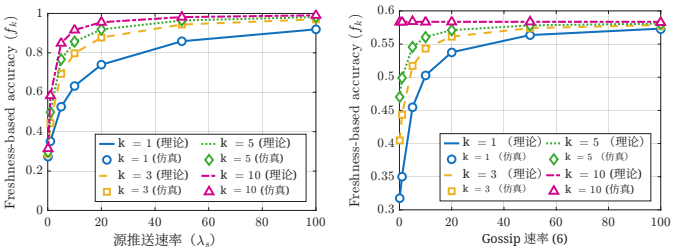


图4. $k \in \{1, 3, 5, 10\}$ 的基于新鲜度的准确度：(a) 与 λ_s 的关系；(b) 与 λ 的关系。

节点行为相同，并且对于所有子集基数$k \in \{1, 2, \cdots, n\}$，平均准确度等于单节点子集的平均准确度。在此，我们还看到，与 λ相比，λ_s 在提高平均准确度方面有效得多。

B. 一般多状态信息源的结果

在第三组采用一般信息源模型的仿真结果中，我们考虑一个具有 $M = 4$ 个状态的连续时间马尔可夫链源，其生成矩阵为

$$\mathbf{Q} = \begin{bmatrix} -1.00 & 0.25 & 0.15 & 0.60 \\ 0.40 & -0.85 & 0.30 & 0.15 \\ 0.20 & 0.85 & -1.15 & 0.10 \\ 0.15 & 0.25 & 0.45 & -0.85 \end{bmatrix}. \tag{88}$$

与二元连续时间马尔可夫链情形一样，网络规模取为 $n = 10$，并使用总计 2,250,000 个时间步来仿真稳态条件。

图4(a)和图4(b)展示了一般多状态信息源（具有4种状态）的仿真结果。随着 λ 增大，每个 k 值对应的平均准确度都收敛到相同的值，这突显了在较高 Gossip 速率下网络表现得如同单个节点。一个具有代表性的例子是，$k = n$ 的平均准确度不会随 λ 变化，而 $k \neq n$ 的所有其他 f_k 值都趋近于f_n。随着我们增大 λ_s，网络中的每个节点都能够获取最新鲜的信息，因此，如图4(a)所示，它们的准确度会越来越接近1。由此，针对 $M = 2$ 状态源的观察结果也适用于一般M状态源。

我们还给出了引理 1 中所得表达式的仿真结果。引理 1 指出，在稳态条件下，持有信息的节点预期比例 q 对于网络中任意基数为 k 的子集 $\pi_q k$ 为给定值。对于图5中针对子集基数 k 的扫描，源推送速率和 Gossip 速率均固定为 1，

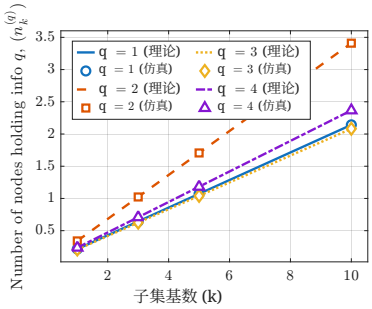


图5. 持有信息 q 的节点数与集合基数 k 的关系（四状态连续时间马尔可夫链信息源）

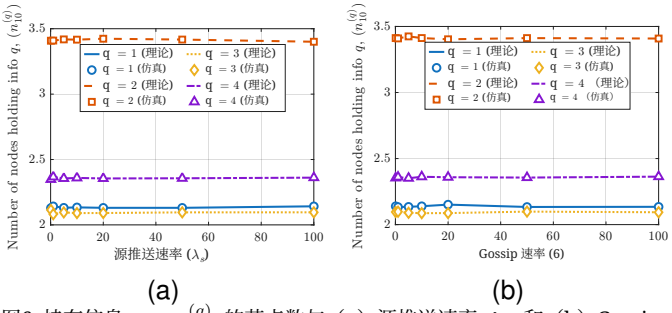


图6. 持有信息 q、$n_k^{(q)}$ 的节点数与（a）源推送速率 λ_s 和（b）Gossip 速率 λ的关系。

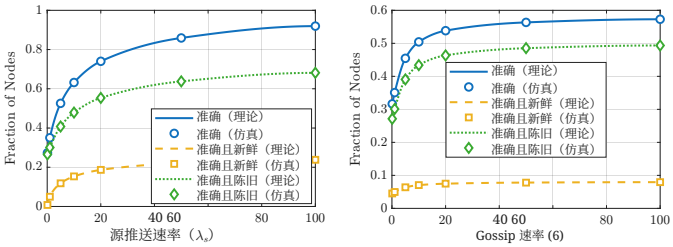
$\lambda_s = \lambda = 1$。对于 $n_k^{(q)}$ 针对源推送速率和 Gossip 速率的扫描，未被扫描的参数固定为 1，即图6(a)中为 $\lambda = 1$，图6(b)中为 $\lambda_s = 1$。正如预期，持有信息的节点数 q随子集基数 k线性增加，如图5所示。此外，图6(a)和(b)展示了整个网络中持有信息的节点数 q，即 $n_n^{(q)}$，在 λ_s 和 λ 分别变化时保持不变。

C. 新鲜度导致的信息准确性结果

在最后一组仿真结果中，我们考察具有准确且严格新鲜信息的节点比例（$G_{k=n}$），并观察近乎瞬时的信息传播速率对准确度的影响。与之前的仿真结果一样，网络规模设为 $n = 10$，而未进行扫描的参数，即图7(a)中的 λ 和图7(b)中的 λ_s，固定为1。注意，$G_k \neq G_j$，$k \neq j$，$k, j \in \mathcal{N}$，因此我们不能任意选择 k 值。由于我们希望观察整个网络的行为，因此在两幅图中都选择 $k = n$。在准确且新鲜的节点比例方面，图7(a)与图7(b)之间存在明显差距。即使 Gossip 速率非常快，也只有 ≈ 0.08 的节点同时准确且严格新鲜，而准确节点（新鲜或陈旧）的总比例要大得多（≈ 0.57）。因此，较高的 λ 主要使节点同步到相同的信息，但并不能保证该信息是当前的；也就是说，新鲜度最终受源推送速率 λ_s 的瓶颈限制。随着 Gossip 速率增大，准确且陈旧的节点比例明显上升，这一点也可以观察到。

VII. 结论与未来工作

本文研究了基于时效性的全连接流言网络中的信息准确性。在这类网络中，节点仅根据信息新鲜度接收数据包



(a) (b)

图7. 整个网络中持有新鲜且准确的信息的节点比与（a）源推送速率 λ_s 和（b）信息传播速率 λ 的关系。

（信息年龄）。根据这一规则，节点可能持有最新鲜的数据包，却仍与源状态不一致。为量化这种不一致，我们引入了平均准确度和基于新鲜度的准确度，并针对二元连续时间马尔可夫链信源开展了基于随机混合系统的稳态分析。我们推导出一种矩阵值向后递推关系，用于刻画带模式标记的基于新鲜度的准确度，并给出了显式边界解，从而能够计算所有子集大小对应的结果。我们通过联合连续时间马尔可夫链方法将这一框架扩展到一般的 M 状态信源，并量化了直接源推送和点对点流言传播的贡献。数值结果和渐近结果表明，提高源推送速率 λ_s 比提高信息传播速率 λ 更能有效提升稳态准确度。未来的工作包括将该框架扩展到一般网络拓扑、多个源过程、最优速率分配，以及适用于更广泛信息传播场景的广义接收规则。

参考文献

[1] Y. Wang、G. Tan、S. Zhou、D. Yang 和 B. Ni，“车联网交叉路口信息年龄的时空分析”，载于 *IEEE HPCC/DSS/SmartCity/DependSys*。IEEE，2021年12月，第901–907页。[2] S. K. Kaul、M. Gruteser、V. Rai 和 J. Kenney，“最小化车联网中的信息年龄”，载于 *IEEE Infocom*，2011年3月。[3] K. Akkaya 和 M. Younis，“无线传感器网络路由协议综述”，*Ad Hoc Networks*，第3卷，第325–349页，2005年5月。[4] L. Villalba、A. Orozco、A. Cabrera 和 C. Abbas，“无线传感器网络中的路由协议”，*Sensors*，第9卷，第8399–8421页，2009年10月。[5] S. He、H. Shin、S. Xu 和 A. Tsourdos，“低成本传感器网络上的分布式估计：最新研究综述”，*Information Fusion*，第54卷，第21–43页，2020年2月。[6] R. D. Yates 和 S. K. Kaul，“信息年龄：多个信源的实时状态更新”，*IEEE*信息论汇刊，第65卷，第1807–1827页，2019年3月。[7] R. D. Yates、Y. Sun、D. R. Brown、S. K. Kaul、E. Modiano 和 S. Ulukus，“信息年龄：介绍与综述”，*IEEE*通信领域精选期刊，第39卷，第1183–1210页，2020年5月。[8] Y. Sun、I. Kadota、R. Talak 和 E. Modiano，《信息年龄：信息新鲜度的新度量》。Springer Nature，2022年。[9] M. Salimnejad、M. Kountouris、A. Ephremides 和 N. Pappas，“信息年龄版本：马尔可夫信源监测的语义视角”，*IEEE*通信汇刊，第73卷，第12期，第14 486–14 502页，2025年12月。[10] A. Maatouk、S. Kriouile、M. Assaad 和 A. Ephremides，“错误信息年龄：一种新的状态更新性能指标”，*IEEE/ACM*网络汇刊，第28卷，第5期，第2215–2228页，2020年10月。[11] I. Cosandal、N. Akar 和 S. Ulukus，“连续时间马尔可夫链信息源的多阈值AoII最优采样策略”，可在 *arXiv* 获取：*2407.08592*，2024年7月。[12] I. Cosandal、S. Ulukus 和 N. Akar，“最小化函数的错误信息年龄以实现远程估计”，可在 arXiv 获取：2504.10451，2025年4月。[13] S. Boyd、A. Ghosh、B. Prabhakar 和 D. Shah，“谣言传播算法：设计、分析与应用”，载于 IEEE年度联合会议：Infocom，IEEE计算机与通信学会，2005年3月。[14] D. Shah，“谣言传播算法”，《网络领域基础与趋势》，第3卷，第1–125页，2008年。[15] S. Deb、M. Medard 和 C. Choute，“代数谣言传播：实现最优多重谣言传播的网络编码方法”，IEEE信息论汇刊，第52卷，第2486–2507页，2006年6月。[16] S. Sanghavi 、 B. Hajek 和 L. Massoulie，“ 传播多条消息 ”，IEEE信息论汇刊，第53卷，第4640–4654页，2007年12月。[17] R. D. Yates，“及时谣言传播”，载于2021年IEEE第22届无线通信信号处理进展国际研讨会（SPAWC），2021年9月，第331–335页。[18] M. A. Abd-Elmagid 和 H. S. Dhillon，“网络中谣言传播年龄的分布”，《熵》，第25卷，第2期，2023年2月。[19] A. Srivastava 和 S. Ulukus，“网格上的谣言传播年龄”，载于2023年第59届Allerton通信、控制与计算年度会议（Allerton），2023年9月，第1–8页。[20] P. Kaswan 和 S. Ulukus，“及时谣言传播网络中的信息突变与错误信息扩散”，载于GLOBECOM 2023—2023年IEEE全球通信大会，2023年12月，第3560–3565页。[21] R. D. Yates，“网络中的谣言传播年龄”，载于2021年IEEE国际信息论研讨会（ISIT），2021年7月，第2984–2989页。[22] P. Kaswan 和 S. Ulukus，“谣言传播年龄对时间戳篡改的敏感性”，载于2022年IEEE信息论研讨会（ITW），2022年11月，第398–403页。[23] B. Buyukates、M. Bastopcu 和 S. Ulukus，“集群谣言传播网络中的信息版本年龄”，IEEE信息论选题领域期刊，第3卷，第1期，第85–97页，2022年3月。[24] P. Kaswan 和 S. Ulukus，“谣言传播网络中的错误信息扩散：传输突变的影响”，IEEE通信汇刊，第73卷，第9期，第7711–7723页，2025年9月。[25] P. Mitra 和 S. Ulukus，“稠密网络中的及时机会式谣言传播”，载于IEEE INFOCOM 2023—IEEE计算机通信会议研讨会（INFOCOM WKSHPS），2023年5月，第1–6页。[26] M. Bastopcu、B. Buyukates 和 S. Ulukus，“采用二元新鲜度度量的谣言传播”，载于2021年IEEE Globecom研讨会（GC Wkshps），2021年12月，第1–6页。[27] E. Delfani 和 N. Pappas，“具有能量采集传感器的谣言传播网络中的版本年龄最优缓存状态更新”，IEEE通信汇刊，第73卷，第4期，第2344–2360页，2025年4月。[28] T. J. Maranzatto、A. Srivastava 和 S. Ulukus，“谣言传播网络中的信息退化与错误信息”，可在 arXiv 获取：2501.13086，2025年1月。[29] J. Luo 和 N. Pappas，“马尔可夫源远程估计中漏报和误报年龄的最小化”，载于 ACM MobiHoc，2024年，第381–386页。[30] ——，“信息年龄与信息语义在马尔可夫源远程估计中的作用”，可在 arXiv 获取：2507.18514，2025年7月。[31] ——，“利用离散状态马尔可夫源远程估计中的数据重要性”，可在 arXiv 获取：2406.18270，2024年6月。[32] K. Saurav、N. B. Shroff 和 Y. Liang，“以采样成本监测马尔可夫系统中的状态转移”，可在 arXiv 获取：2510.22327，2025年10月。[33] P. Kaswan、P. Mitra、A. Srivastava 和 S. Ulukus，“谣言传播网络中的信息年龄：通俗入门与文献综述”，IEEE通信汇刊，第73卷，第8期，第6200–6220页，2025年8月。[34] B. Abolhassani、J. Tadrous、A. Eryilmaz 和 E. Yeh，“动态内容的新鲜缓存”，载于IEEE INFOCOM 2021—IEEE计算机通信会议，2021年5月，第1–10页。[35] J. P. Hespanha，“随机混合系统的建模与分析”，《IEE学报：控制理论与应用》，第153卷，第5期，第520–535页，2006年9月。[36] R. D. Yates，“网络中的信息年龄：矩、分布与采样”，IEEE信息论汇刊，第66卷，第9期，第5712–5728页，2020年9月。[37] N. Akar 和 S. Ulukus，“异质连续时间马尔可夫链的基于查询的采样：采用二元新鲜度的建模与优化”，IEEE通信汇刊，第72卷，第12期，第7705–7714页，2024年。[38] S. Liyanaarachchi、S. Ulukus 和 N. Akar，“利用感知年龄最大化基于查询的更新系统中的新鲜度”，可在 arXiv 获取：2601.14075，2026年1月。