Report 1: Information Freshness Optimization in Graphs via Reinforcement Learning

Roman Iashchenko
HSE University
Email: roma.yashchenko3@gmail.com

Manyou Ma (Instructor)
School of Artificial Intelligence,
Shenzhen Technology University, Shenzhen, China
Email: mamanyou@sztu.edu.cn

***Abstract*—We study how to keep information *fresh* on a network when only one node can be actively refreshed at a time. We model the problem as a Markov Decision Process (MDP) over graphs where the state is the vector of node ages (AoI), the action is the choice of a node to refresh, and the reward penalizes average age and refresh cost. Updates can probabilistically propagate along edges, reducing neighbors' ages. We motivate the task, situate it within Age of Information (AoI) literature, and give a precise problem formulation suitable for deep reinforcement learning (e.g., DQN) with clear baselines.**

I. Motivation and Background

Timely data is critical in sensing/IoT networks, social platforms, content delivery, and telemetry. The right abstraction for timeliness is *Age of Information* (AoI), the time elapsed since the newest received update was generated [1], [2]. AoI departs from classical delay/throughput metrics and has driven new scheduling and queuing results [3]–[7].

In many real systems, resources are limited: we cannot refresh all nodes in each time step. Moreover, updates often diffuse (e.g., via gossip or cache coherence), making the "where to refresh next" decision a global control problem over a graph. Hand-crafted policies (e.g., "refresh the stalest node") can be brittle under topology changes, heterogeneous link reliabilities, or correlated sources. Reinforcement learning (RL) offers a data-driven alternative that can adapt to dynamics and exploit structure [8]–[11].

Our objective is a compact, CPU-friendly simulator (about 10 nodes) that compares intuitive heuristics to a learned policy and demonstrates reduced average AoI in finite-horizon episodes.

II. Related Work

AoI foundations. The AoI metric and its implications were introduced and developed in seminal works on status updating and queues: Kaul *et al.* analyze update rates and AoI [3], and status updates through queues with LCFS/FCFS disciplines [4]. Sun *et al.* formalize optimal update timing ("update or wait") with age-penalty functions [2]. Yates provides distributional analyses for networks [5]. A broad survey consolidates AoI models and methods [1].

Scheduling and scaling. Kadota *et al.* derive performance guarantees for low-complexity policies (e.g., Max-Weight, Whittle) in unreliable wireless networks [6]. Buyukates *et al.* study AoI scaling in large networks [7].

RL for AoI. RL has been used to minimize AoI under HARQ and resource constraints [8]; deep RL to control UAV-assisted status collection [9]; and DRL to schedule correlated sources via AoCI [10]. AoI-aware resource scheduling with DRL in IIoT is also explored [11]. Recent work extends to correlated-source settings and mobility, motivating graph-aware policies [12].

Positioning. We focus on *graph-structured* freshness control with simple diffusion. Our setup is intentionally minimal to fit a two-page report and a small Colab simulation while remaining faithful to the AoI literature.

III. System Model and Problem Formulation

A. Graph and Age Dynamics

Let $G = (V, E)$ be an undirected simple graph with $|V| = N$. At discrete time $t = 0, 1, 2, \ldots$, each node i has an *age* $a_i(t) \in \mathbb{Z}_{\geq 0}$. Ages increase by one each step unless reset by a refreshed or propagated update.

At time t, an agent selects a node $u_t \in V$ to *refresh*. Define a per-edge propagation probability $p \in (0, 1]$ and let an instantaneous independent propagation process attempt to transmit the new update from u_t to each neighbor $j \in \mathcal{N}(u_t)$. On success, j's age resets to $\min\{a_j(t^+), a_{u_t}(t^+)\}$; in our canonical model, a direct refresh sets $a_{u_t}(t^+) = 0$, and any successful propagation resets affected neighbors to 0 as well. (Multi-hop propagation within the same step can be disabled for simplicity; we use 1-hop as default to maintain Markovian and local transitions.)[1]

B. MDP Specification

State. $s_t = (a_1(t), \ldots, a_N(t)) \in \mathbb{Z}^N_{\geq 0}$. Optional features (not required): degree vector, last refreshed node id.

Action. $a_t \in \{1, \ldots, N\}$ selects which node to refresh. (Optional ablations: a "wait" action or a "boosted refresh" with higher p and explicit cost.)

Transition. Deterministic age increment by $+1$ for all nodes, then stochastic resets at u_t and (independently) at neighbors via Bernoulli(p). This yields a well-defined controlled Markov chain on $\mathbb{Z}^N_{\geq 0}$.

[1] Keeping 1-hop propagation with probability p mirrors widely used diffusion/gossip abstractions while preserving a compact transition model.

Reward. We target low average AoI with a small refresh penalty:

$$r_t = -\frac{1}{N}\sum_{i=1}^{N} a_i(t^+) \; - \; \lambda \cdot \underbrace{\mathbb{1}\{a_t \neq \text{wait}\}}_{\text{refresh cost}}, \qquad (1)$$

with $\lambda \geq 0$. Episodes have horizon T, objective $\max_\pi \mathbb{E}_\pi\left[\sum_{t=0}^{T-1} r_t\right]$.

Baselines. (i) *Random*: pick a node uniformly. (ii) *Greedy-stale*: pick $\arg\max_i a_i(t)$. (iii) *Degree-weighted*: pick node maximizing $a_i(t) \cdot \deg(i)$ to exploit diffusion potential. These echo AoI scheduling intuitions [6], [7].

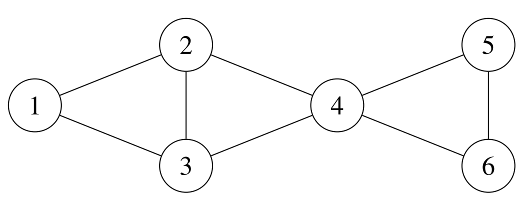
RL Agent. For a small N (e.g., $N{=}10$), a DQN with a multi-layer perceptron over the age vector suffices; the N-way discrete action head outputs Q-values for nodes. Exploration via ϵ-greedy and a replay buffer; reward normalization stabilizes training [8], [10].

C. Environment Step (Pseudo-code)

Algorithm 1 Environment step at time t

Input: $a(t) \in \mathbb{Z}_{\geq 0}^N$, action $u \in V$, graph G, $p \in (0,1]$
$a \leftarrow a(t) + \mathbf{1}$ ▷ all ages increase by 1
$a_u \leftarrow 0$ ▷ refresh chosen node
for $j \in \mathcal{N}(u)$ **do**
    With prob. p, set $a_j \leftarrow 0$
end for
$r \leftarrow -\frac{1}{N}\sum_{i=1}^{N} a_i \; - \; \lambda$
return (a, r)

D. Illustration



Example graph ($N{=}6$). Refreshing node 4 resets $a_4 \leftarrow 0$; each neighbor (2,3,5,6) resets to 0 independently with probability p.

Fig. 1: One-step refresh with 1-hop probabilistic propagation.

IV. Notes on Assumptions and Scope

We adopt the standard AoI definition and a minimal stochastic diffusion rule consistent with gossip-like propagation. Our evaluation will (i) report average AoI over episodes, (ii) compare against random/greedy baselines, and (iii) perform sensitivity to p, λ, and topology. This keeps the project faithful to AoI literature while remaining computationally simple.

Acknowledgment

The topic aligns with the RL course requirements (states/actions/rewards ≥ thresholds) and is grounded in peer-reviewed AoI research.

References

[1] R. D. Yates, Y. Sun, D. R. Brown III, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of Information: An Introduction and Survey," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1183–1210, May 2021.
[2] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or Wait: How to Keep Your Data Fresh," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7492–7508, Nov. 2017.
[3] S. K. Kaul, R. D. Yates, and M. Gruteser, "Real-time status: How often should one update?," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 2731–2735.
[4] S. K. Kaul, R. D. Yates, and M. Gruteser, "Status updates through queues," in *Proc. 46th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2012, pp. 1–6.
[5] R. D. Yates, "The Age of Information in Networks: Moments, Distributions, and Sampling," *IEEE Trans. Inf. Theory*, vol. 66, no. 9, pp. 5712–5728, Sept. 2020.
[6] I. Kadota, A. Sinha, and E. Modiano, "Scheduling Policies for Minimizing Age of Information in Wireless Networks," *IEEE/ACM Trans. Netw.*, vol. 26, no. 6, pp. 2637–2650, Dec. 2018.
[7] B. Buyukates, A. Soysal, and S. Ulukus, "Age of Information Scaling in Large Networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.
[8] E. T. Ceran, D. Gündüz, and A. György, "A Reinforcement Learning Approach to Age of Information in Multi-User Networks with HARQ," *arXiv:2102.09774*, 2021.
[9] M. A. Abd-Elmagid, A. Ferdowsi, H. S. Dhillon, and W. Saad, "Deep Reinforcement Learning for Minimizing Age-of-Information in UAV-Assisted Networks," in *Proc. IEEE GLOBECOM*, Dec. 2019, pp. 1–6.
[10] B. Yin, S. Zhang, and Y. Cheng, "Application-Oriented Scheduling for Optimizing the Age of Correlated Information: A Deep-Reinforcement-Learning-Based Approach," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8748–8759, Sept. 2020.
[11] H. Li, L. Tang, S. Chen, *et al.*, "AoI-Aware Resource Scheduling for Industrial IoT with Deep Reinforcement Learning," *Electronics*, vol. 13, no. 6, art. 1104, Mar. 2024.
[12] C. Xu, Z. Wang, and H. Wang, "Optimal Status Updates for Minimizing Age of Correlated Information," *IEEE Trans. Mobile Comput.*, vol. 23, no. 12, pp. 6848–6864, Dec. 2024.