SCIENCE CHINA
Information Sciences

• LETTER •

September 2026, Vol. 69, Iss. 9, 199204:1–199204:2
https://doi.org/10.1007/s11432-025-4899-1

Pinning stabilization of Boolean networks via reinforcement learning

Yanxin FAN[1], Amol YERUDKAR[2*], Yang LIU[1,3,4*] & Weihua GUI[5]

[1]*School of Mathematical Sciences, Zhejiang Normal University, Jinhua 321004, China*
[2]*School of Computer Science and Technology, Zhejiang Normal University, Jinhua 321004, China*
[3]*China-Mozambique Belt and Road Joint Laboratory on Smart Agriculture, Zhejiang Normal University, Jinhua 321004, China*
[4]*Hangzhou School of Automation, Zhejiang Normal University, Hangzhou 311231, China*
[5]*School of Automation, Central South University, Changsha 410083, China*

Received 27 July 2025/Revised 23 October 2025/Accepted 25 February 2026/Published online 1 September 2026

Citation Fan Y X, Yerudkar A, Liu Y, et al. Pinning stabilization of Boolean networks via reinforcement learning. Sci China Inf Sci, 2026, 69(9): 199204, https://doi.org/10.1007/s11432-025-4899-1

Introduced by Kauffman [1], Boolean networks (BNs) provide an efficient computational framework for modeling gene regulatory networks (GRNs) using binary states and Boolean functions. These networks are effective in capturing basic biological dynamics. Research on BNs often employs the semi-tensor product (STP) framework [2], which facilitates the algebraic formalization of logical dynamics. When external control is applied, BNs become Boolean control networks (BCNs), where stabilization is a key challenge. Pinning control strategically manipulates a subset of nodes rather than all nodes to achieve global stabilization, reducing control costs. In order to control biological networks, the computational complexity under the STP framework may be high. Furthermore, it is difficult to fully grasp the dynamics of GRNs. Reinforcement learning (RL) [3] provides a potential solution. A BN is a discrete dynamical system. As a standard reinforcement learning method, the Q-learning (QL) algorithm explores discrete action spaces to identify optimal control policies. Most existing RL applications do not explicitly address the problem of pinning control or the identification of the minimal pinned node set.

In this article, we apply the QL algorithm to address the pinning control problem of BNs with unknown network structure. The main contributions of this work are as follows. (i) The process of identifying pinned nodes is formalized through a two-phase algorithm based on attractor basin analysis and Hamming distance computation between network states. (ii) A model-free reinforcement learning framework for the pinning control of BNs is established under the Markov decision process (MDP) formulation. (iii) Within this framework, a QL algorithm is introduced to simultaneously determine the minimal pinned node set and the state feedback matrix. Through this design, global stabilization is achieved without prior knowledge of the network structure. For completeness, more detailed discussions on the background and preliminaries are included in Appendix A.

Consider a BN with m designated pinned nodes indexed by $\mathcal{I} = \{1, 2, \ldots, m\}$. Under the pinning control strategy, the original BN transforms into a BCN as follows:

$$\begin{aligned} x_i(t+1) &= \hat{f}_i(c_i(t), x_1(t), x_2(t), \ldots, x_n(t)), i \in \mathcal{I}, \\ x_j(t+1) &= f_j(x_1(t), x_2(t), \ldots, x_n(t)), j \in [1, n]\backslash\mathcal{I}, \end{aligned} \tag{1}$$

where $\hat{f}_i, i \in \mathcal{I}$ are modified Boolean functions, and $c_i, i \in \mathcal{I}$ are feedback controls applied to the i-th pinned node.

Definition 1. Consider all pinned nodes as a set. The collection of controlled nodes constitutes the pinned set $\Omega := \{x_i\}_{i=1}^m \subseteq V$, where $V = \{x_1, \ldots, x_n\}$ enumerates all network nodes.

Let the state and input control be denoted by $\mathbf{X} = (x_1, \ldots, x_n) \in \mathcal{D}^n$ and $\mathbf{C} = (c_1, \ldots, c_m) \in \mathcal{D}^m$, respectively. Let $\Phi(t; \mathbf{X}_0, \mathbf{C}) := \mathbf{X}(t)$ denote the trajectory satisfying the initial condition $\Phi(0; \mathbf{X}_0, \mathbf{C}) := \mathbf{X}_0$. Then we review the global stabilizability of BCNs.

Definition 2 (See [2])**.** A BCN is globally stabilizable to state $\mathbf{X}^*$ if for every initial state $\mathbf{X}_0 \in \mathcal{D}^n$, there exists a control sequence $\mathbf{C} \in \mathcal{D}^m$ and a finite time $\tau \in \mathbb{Z}^+$ such that $\mathbf{X}(t) = \mathbf{X}^*$, $\forall t \geqslant \tau$.

Identification of pinned nodes. Consider a BCN with state space $\mathcal{X} = \mathcal{D}^n$. The attractor $\mathcal{G} \subset \mathcal{X}$ is a set to which the initial state trajectory converges, and the i-th attractor in it is denoted by $\mathcal{G}_i$. Its basin $\mathcal{B}(\mathcal{G}_i)$ contains all states $\mathbf{X} \in \mathcal{X}$ satisfying $\lim_{t\to\infty} \Phi(t; \mathbf{X}, \mathbf{C}) = \mathcal{G}_i$. The identification of pinned nodes in a BN necessitates characterization of attractors and their associated basins. Identifying pinned nodes is achieved by comparing distances of relevant states.

Definition 3. Given two states $\mathbf{X}_a = [x_a^1, x_a^2, \ldots, x_a^n]$ and $\mathbf{X}_b = [x_b^1, x_b^2, \ldots, x_b^n]$ in a BN, their Hamming distance is defined as

$$\mathbf{d}(\mathbf{X}_a, \mathbf{X}_b) = \sum_{i=1}^{n} |x_a^i - x_b^i|. \tag{2}$$

This indicator quantifies the difference in node activation patterns between two states.

* Corresponding author (email: ayerudkar@zjnu.edu.cn, liuyang@zjnu.edu.cn)

Definition 4. For any pair of states $\mathbf{X}_a$ and $\mathbf{X}_b$, the differential node set is defined as

$$\Xi(\mathbf{X}_a, \mathbf{X}_b) = \{x^j | j \in \{1, 2, \dots, n\}, |x_a^j - x_b^j| > 0\}. \quad (3)$$

This set identifies critical nodes that cause state differences.

We identify the pinned set Ω in two phases. In the first phase, we focus on the desired state $\mathbf{X}_r$ that will become the final attractor. The initial pinned set is determined by comparing the distance between $\mathbf{X}_r$ and its next state. In the second phase, we recursively add nodes to the initial pinned set so that all states of the network can reach the desired state.

Phase 1: Initial pinned set identification

Lemma 1. Let $\mathbf{X}_r$ denote the desired attractor and $\bar{\mathbf{X}}_r$ denote its subsequent state. For any node i satisfying $\mathbf{X}_r^i \neq \bar{\mathbf{X}}_r^i$, i must belong to the pinned set Ω.

According to Lemma 1, the initial pinning set is constructed as $\Omega = \Xi(\mathbf{X}_r, \bar{\mathbf{X}}_r)$. The basin of the desired attractor $\mathbf{X}_r$, denoted by $\mathcal{B}(\mathbf{X}_r)$, can be determined.

Phase 2: Recursive pinned set expansion

To achieve global stabilization, all original attractors in $\mathcal{G}$ must be progressively steered to $\mathcal{B}(\mathbf{X}_r)$. First, we define the minimal basin distance.

Definition 5. Given a state $\mathbf{X}_k$ and a basin set $\mathcal{B}(\mathbf{X}_r)$, the minimal distance between them is defined as

$$\mathrm{d}_{\mathcal{B}}(\mathbf{X}_k) = \min_{\mathbf{X} \in \mathcal{B}(\mathbf{X}_r)} \mathrm{d}(\mathbf{X}_k, \mathbf{X}). \quad (4)$$

By solving the optimization problem:

$$\mathbf{X}^* = \arg\min_{\mathbf{X} \in \mathcal{B}(\mathbf{X}_r)} \mathrm{d}(\mathcal{G}_i, \mathbf{X}), \quad (5)$$

where $\mathcal{G}_i$ is the i-th attractor in the attractor set $\mathcal{G}$. We identify the nearest neighbor state $\mathbf{X}^*$ and expand the pinned set: $\Omega \leftarrow \Omega \cup \Xi(\mathcal{G}_i, \mathbf{X}^*)$. By performing optimal analysis on each attractor in $\mathcal{G}$, the final pinned set can be obtained recursively. The iterative process is described in Appendix B.

Theorem 1. The pinned set generated by the two-phase method is minimal.

The proof of the theorem can be found in Appendix C.

QL for pinning control. Controlling large-scale GRNs with STP-based methods may face the *curse of dimensionality.* In contrast, reinforcement learning methods are competitive as they can be applied in model-free environments. Within the QL framework, we formulate a feedback stabilization approach for BNs and introduce a QL-based algorithm to identify pinned nodes while constructing state feedback control laws.

Consider the following form of feedback law:

$$c_i(t) = k_i(x_1(t), x_2(t), \dots, x_n(t)), \quad (6)$$

where $i \in [1, m]$, k_i is a Boolean function. We can also obtain the algebraic representation by using STP: $\mathbf{c}(t) = \mathbf{K}\mathbf{x}(t)$. $\mathbf{K}$ is called the state feedback matrix. We construct its algebraic representation instead of directly designing a state feedback controller.

In the process of identifying pinned nodes, we stabilize the entire network by incorporating unstable states into the basin of the desired attractor. When some edges in the state transition graph are changed, the node values of these states are essentially modified. Consequently, the control strategy for each state can be obtained. Assume that Algorithm B1 (see Appendix B) provides a pinned set with m nodes, and the deterministic control strategy $\lambda(\mathbf{X}_t, \mathbf{X}_r)$, $\forall \mathbf{X}_t \in \mathcal{D}^n$. Then, upon defining $\lambda_j := \lambda(\mathbf{X}_j, \mathbf{X}_r)$, $j \in [1, 2^n]$, the state feedback matrix, which may not be unique, is represented by $K \in \mathcal{L}^{2^m \times 2^n}$ as $K = [\delta_{2^m}^{\lambda_1} \quad \delta_{2^m}^{\lambda_2} \quad \dots \quad \delta_{2^m}^{\lambda_{2^n}}]$.

Consider the discrete-time system governed by

$$\mathbf{X}_{t+1} = F(\mathbf{C}_t, \mathbf{X}_t), \quad (7)$$

where $F : \mathcal{D}^{n+m} \to \mathcal{D}^n$ is an unspecified function, and $\mathbf{X}_r \in \mathcal{D}^n$ is an equilibrium state.

Let the BCN (1) be modeled as $(\mathcal{D}^n, \mathcal{D}^m, \mathcal{P}, \mathcal{R}, \gamma)$, where $\mathcal{P}$ is unknown to the agent, and $\mathbf{X}_r \in \mathcal{D}^n$ denotes a prescribed equilibrium state. Under appropriate reward function design, the optimal feedback stabilization problem admits the formulation:

$$\max_{\mathbf{C}} \mathbf{E}_{\mathbf{C}} \left[\sum_{t=0}^{\infty} \gamma^t r_{t+1}(\mathbf{X}_t, \mathbf{C}_t, \mathbf{X}_{t+1}, \mathbf{X}_r) \right], \forall \mathbf{X}_0 \in \mathcal{D}^n. \quad (8)$$

The design of the reward function r_{t+1} must align with specific control objectives. For stabilization to $\mathbf{X}_r$ with minimum nodes, we define

$$r_{t+1} = \begin{cases} 10 - c \cdot n_p, & \text{if } \mathbf{X}_{t+1} = \mathbf{X}_r, \\ 0, & \text{otherwise}, \end{cases} \quad (9)$$

where n_p denotes the count of pinned nodes, and $c > 0$ is a tunable penalty parameter.

The l-step controllable set is defined as follows:

$$\Psi_l(\mathbf{X}_r) = \{\mathbf{X}_0 \in \mathcal{D}^n : \exists \mathbf{U}_l \text{ such that } \Phi(l; \mathbf{X}_0, \mathbf{C}_l) = \mathbf{X}_r\}. \quad (10)$$

Theorem 2. Consider a BCN (1) characterized by the MDP quintuple $(\mathcal{D}^n, \mathcal{D}^m, \mathcal{P}, \mathcal{R}, \gamma)$ with the equilibrium state $\mathbf{X}_r \in \mathcal{D}^n$, the system is globally stabilizable to $\mathbf{X}_r$ if and only if there exists an integer $\tau \in \mathbb{Z}^+$ satisfying $\Psi_\tau(\mathbf{X}_r) = \mathcal{D}^n$.

The proof of the theorem can be found in Appendix C.

The feedback stabilization problem (8) is addressed via the QL algorithm, where the Q-values are updated based on the temporal difference (TD) error. The update rule is as follows: $Q_{t+1}(\mathbf{X}_t, \mathbf{C}_t, \mathbf{X}_r) = Q_t(\mathbf{X}_t, \mathbf{C}_t, \mathbf{X}_r) + \alpha_t[TD_{t+1}]$, $TD_{t+1} = r_{t+1} + \gamma \max_{\mathbf{C} \in \mathcal{D}^m} Q_t(\mathbf{X}_{t+1}, \mathbf{C}, \mathbf{X}_r) - Q_t(\mathbf{X}_t, \mathbf{C}_t, \mathbf{X}_r)$. This results in a unique solution for $Q^*(\mathbf{X}_t, \mathbf{C}_t, \mathbf{X}_r)$, leading to the optimal policy: $\lambda^*(\mathbf{X}_t, \mathbf{X}_r) = \arg\max_{\mathbf{C} \in \mathcal{D}^m} Q^*(\mathbf{X}_t, \mathbf{C}, \mathbf{X}_r), \forall \mathbf{X}_t \in D^n$ attainable when Theorem 2 holds during training.

Lemma 2. For BCN (7) satisfying Theorem 2, QL convergences to the fixed point $Q^*(\mathbf{X}_t, \mathbf{C}_t, \mathbf{X}_r), \forall \mathbf{X}_t \in D^n, \forall \mathbf{C}_t \in D^m$, requires the following conditions to be ensured: (i) $\sum_{t=0}^{\infty} \alpha_t = \infty$ and $\sum_{t=0}^{\infty} \alpha_t^2 < \infty$; (ii) variance: $\mathrm{var}[r_t] < \infty$.

Based on Theorem 2, an algorithm of pinning stabilization using QL is presented in Appendix B2.

Acknowledgements This work was supported by National Key Research and Development Program of China (Grant No. 2024YFE0214000), Natural Science Foundation of Zhejiang Province of China (Grant No. LRG25F030002), and Zhejiang Province Leading Geese Plan (Grant No. 2025C01056).

Supporting information Appendixes A–D. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Kauffman S A. Metabolic stability and epigenesis in randomly constructed genetic nets. J Theoret Biol, 1969, 22: 437–467
2 Cheng D Z, Qi H S, Li Z Q. Analysis and Control of Boolean Networks: A Semi-Tensor Product Approach. London: Springer, 2011
3 Sutton R S, Barto A G. Reinforcement Learning: An Introduction. Cambridge: MIT press, 2018