Advanced Topics in Control: Distributed Systems and Control

Lecture 7: Markov Chains and Stochastic Stability

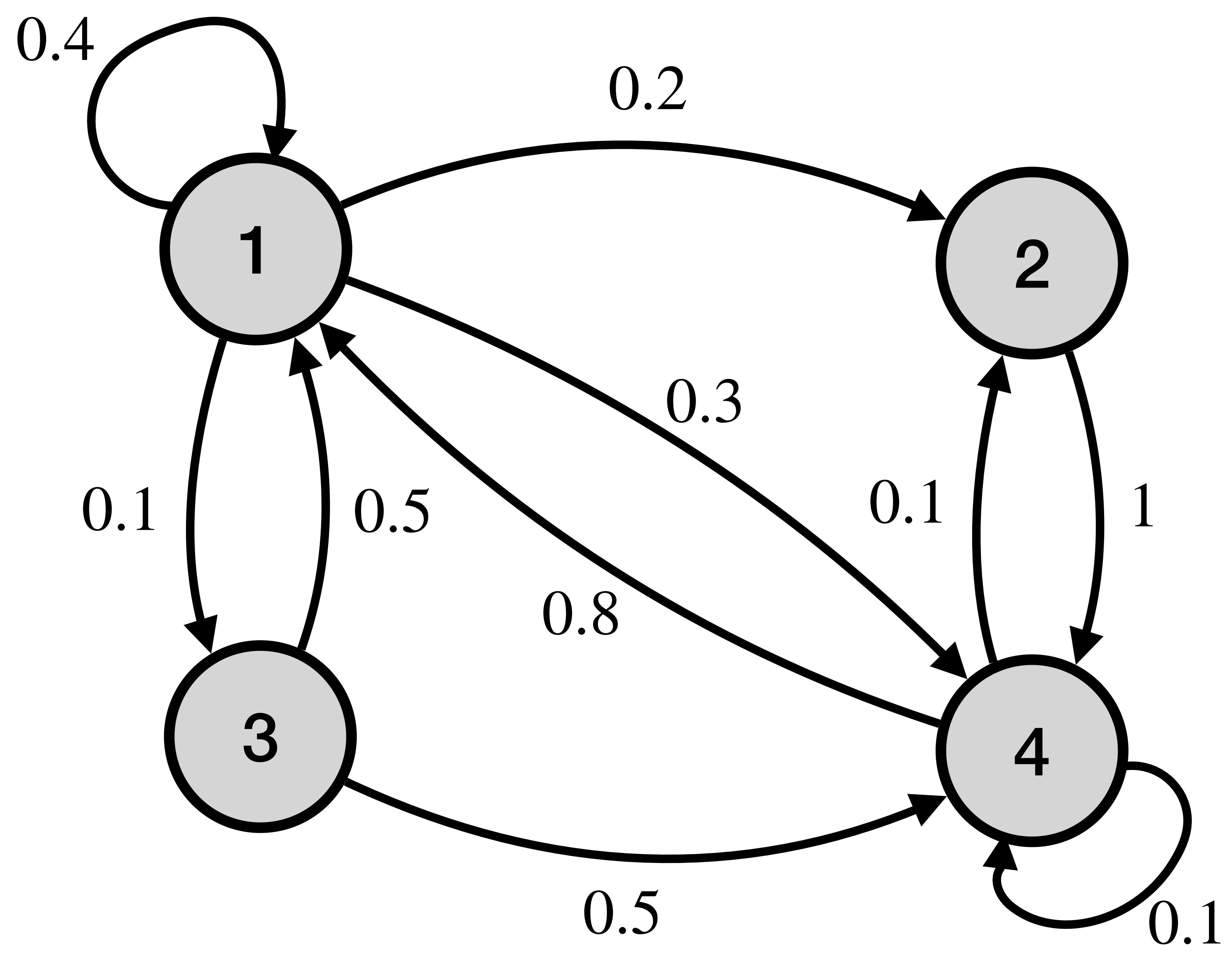
Dr. Mohammad H. (Vahid) Mamduhi, April 4th, 2022

ETH zürich
AUTOMATIC CONTROL LABORATORY

A mind-refreshing example from Lectures 3 & 4

For the given graph $G(\mathscr{V}, \mathscr{E}, \{a_e\}_{e\in\mathscr{E}})$, we learned the followings:

- $\exists$ a directed path from any node in $\mathscr{V}$ to any other node $\rightarrow$ G is strongly connected
- G has self-cycles $\rightarrow$ G is aperiodic
- Out-degree of every node is one $\rightarrow$ adjacency matrix is row-stochastic
- G is strongly connected and aperiodic $\rightarrow$ adjacency matrix is primitive



The representative graph G

Adjacency matrix:

$$A = \begin{pmatrix} 0.4 & 0.2 & 0.1 & 0.3 \\ 0 & 0 & 0 & 1 \\ 0.5 & 0 & 0 & 0.5 \\ 0.8 & 0.1 & 0 & 0.1 \end{pmatrix}$$

Remember this example…!

Graph G represents the discrete time averaging algorithm

$$x(t+1) = Ax(t)$$

☑ **A is primitive $\rightarrow$ convergence to consensus!**

Probability spaces: basic concepts

A probability space is represented by 3 components $(\Omega, \mathscr{A}, \mathrm{P})$:

- sample space Ω: a non-empty set containing all possible outcomes of a process or an experiment
- sigma-algebra $\mathscr{A}$: collection of all subsets of the sample space Ω
- probability measure $\mathrm{P} : \mathscr{A} \to [0,1]$: a measure which assigns a probability to each of the elements/events in $\mathscr{A}$

Countable state spaces

A probability space is called countable if it is discrete and contains a finite or countable number of elements. The sigma-algebra $\mathscr{A}$ is then the set of all subsets of Ω (including $\varnothing$ and Ω)

Example: Let $\Omega = \{a, b, c\}$, the possible outcomes:

- Some possible events: $\{b\}, \{a, b\}, \{b, c\}, \{a, b, c\}, \ldots$
- Collection of all subsets of $\Omega = \{\{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}, \varnothing\} = \mathscr{A}$
- Probability measure is $\mathrm{P} = \{p_{\{a\}}, p_{\{b\}}, p_{\{c\}}, p_{\{a,b\}}, p_{\{a,c\}}, p_{\{b,c\}}, p_\Omega\}$

with $p_\varnothing = 0$ and $p_{\{a\}} + p_{\{b\}} + p_{\{c\}} = p_{\{a,b\}} + p_{\{c\}} = p_{\{a,c\}} + p_{\{b\}} = p_{\{b,c\}} + p_{\{a\}} = p_\Omega = 1$

Probability spaces: basic concepts

Random variable: on a countable space $(\Omega, \mathscr{A}, \mathrm{P})$, the variable $X : \Omega \to \mathbb{R}$ is called a real-valued random variable. Replace $\mathbb{R}$ with $\mathbb{R}^{n\times m}$ and X is a random matrix of size $n \times m$.

Expected value (mean): a random variable $X \in \mathbb{R}$ defined on a countable space with possible outcomes x_i each with the probability of occurrence p_i has the expected value

$$\mathrm{E}[X] = \sum_i x_i \, p_i$$

- For a random matrix $X \in \mathbb{R}^{n\times m}$ defined on a countable set with random elements X_{ij} with expected values $\mathrm{E}[X_{ij}]$ the expected value matrix equals the matrix of its expected elements

$$\mathrm{E}[X] = \Big[\mathrm{E}[X_{ij}]\Big]_{n\times m}$$

- For two random variables $\{X, Y\}$ with finite means, If $X \leq Y$, *almost surely,* then $E[X] \leq E[Y]$

<u>Question</u>: A fair coin has two outcomes "head" and "tail" with the expected values of 0.5. You flip it 100 times, and observe 35 heads and 65 tails. How this can be explained w.r.t. the expected value?

Time-homogeneous Markov chains on countable spaces

Let a stochastic process $\Phi = \{\phi_0, \phi_1, \phi_2, \dots\}$ be defined on a countable space $(\Omega, \mathscr{A}, \mathsf{P})$, where $p^n(x_0, \mathscr{B})$ is the probability the process will be in set $\mathscr{B} \in \mathscr{A}$ after n transitions, given $\phi_0 = x_0$, i.e.,

$$p^n(x_0, \mathscr{B}) \triangleq \mathsf{P}(\phi_n \in \mathscr{B} \,|\, \phi_0 = x_0)$$

1. Φ is a Markov chain if

$$\mathsf{P}(\phi_{n+1} \in \mathscr{B} \,|\, \phi_n = x_n, \dots, \phi_0 = x_0) = \mathsf{P}(\phi_{n+1} = \mathscr{B} \,|\, \phi_n = x_n) = p(x_n, \mathscr{B})$$

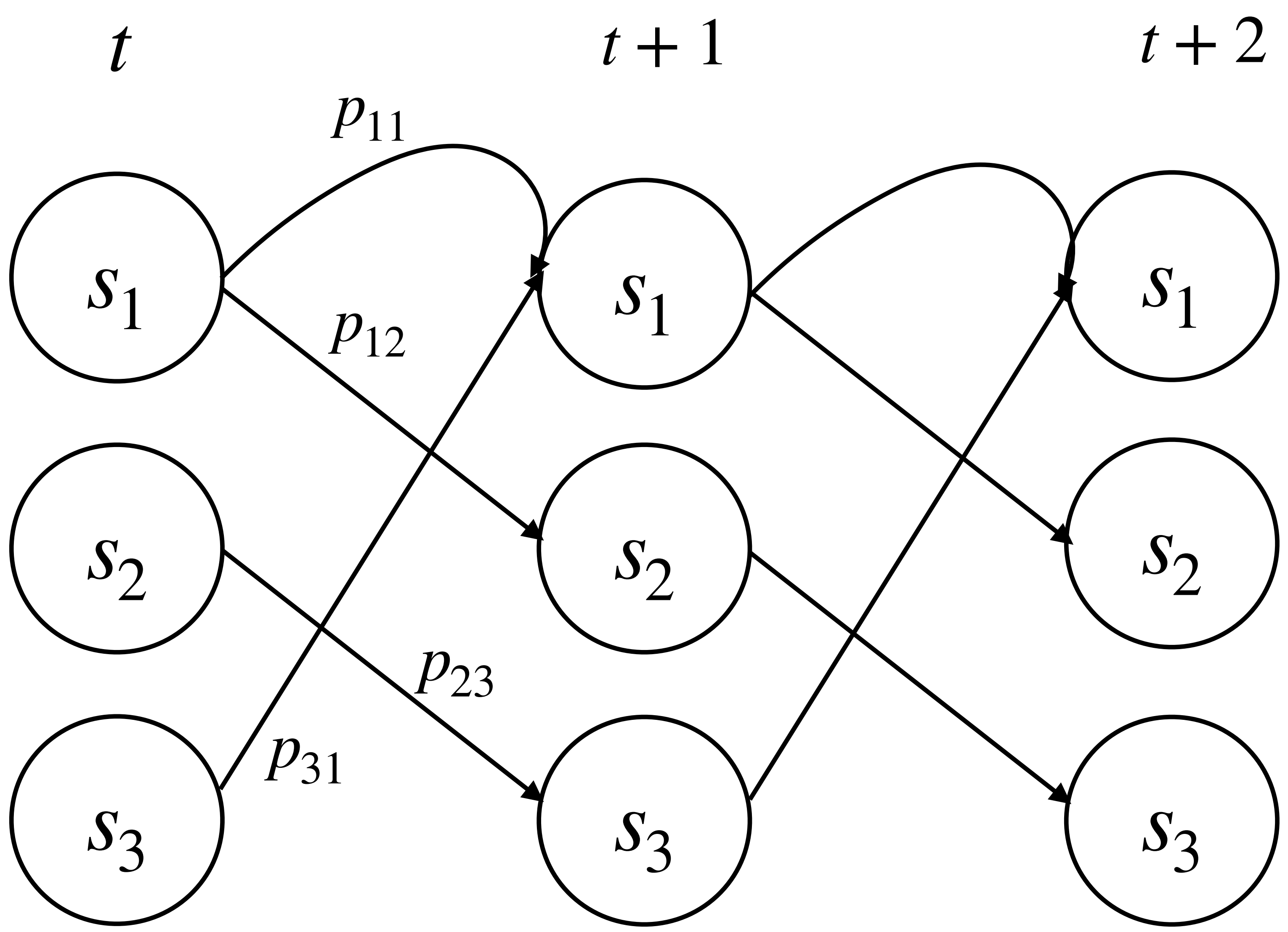
2. Φ is a time-homogeneous Markov chain if

$$\mathsf{P}(\phi_{n+m} \in \mathscr{B} \,|\, \phi_n = x_n, \dots, \phi_0 = x_0) = \mathsf{P}(\phi_{n+m} = \mathscr{B} \,|\, \phi_n = x_n) = p^m(x_n, \mathscr{B})$$

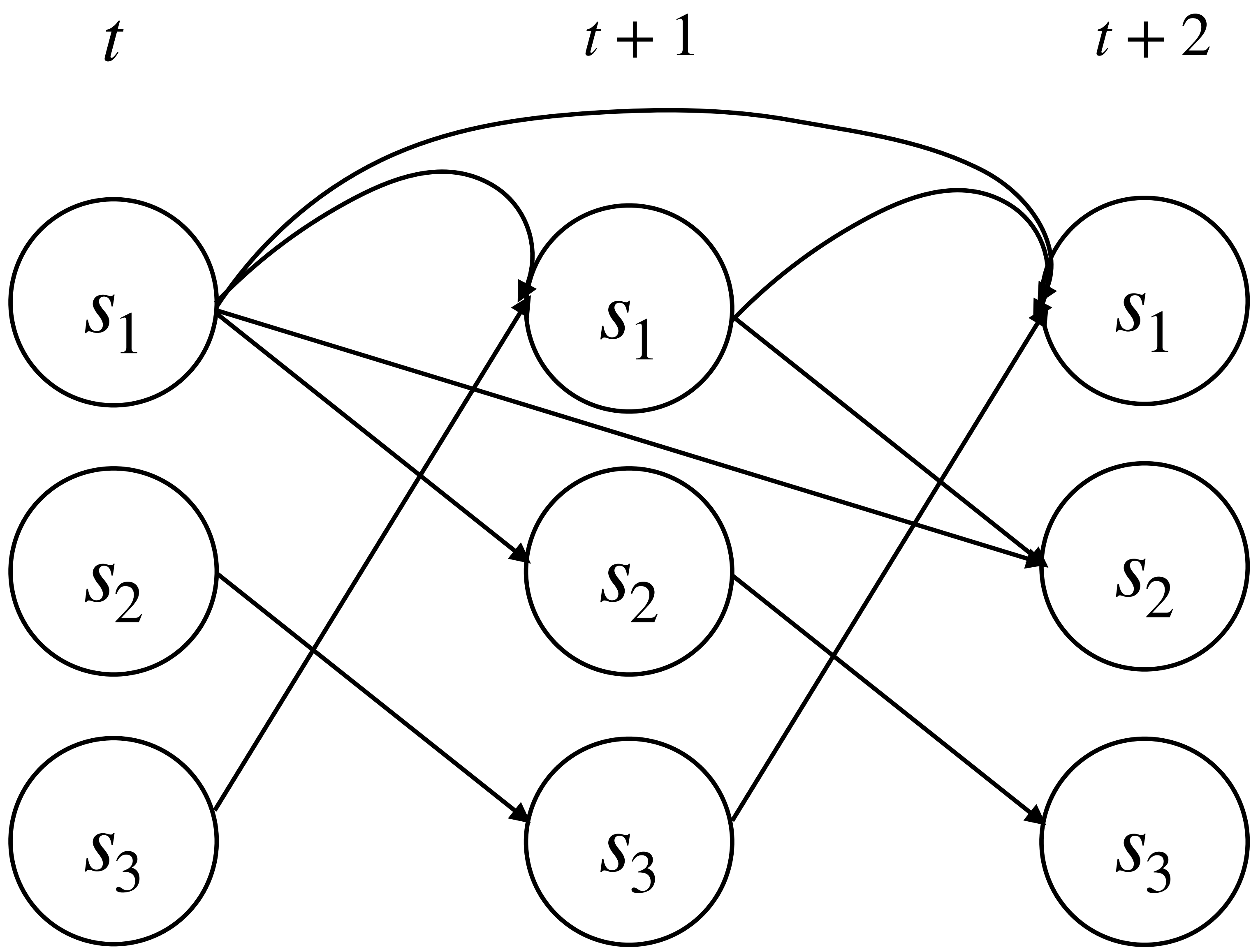
- Markov property: independence of $p^m(x_n, \mathscr{B})$ from $\phi_j, j < n$ (any similarity with $x_{k+1} = Ax_k$?)
- Time-homogeneity property: Independence of $p^m(x_n, \mathscr{B})$ from the index n
- A Markov chain is called finite, if its random states can take values from a finite set

☑ *On finite and countable spaces, $\mathscr{B}$ reduces to a discrete event in $\mathscr{A}$*

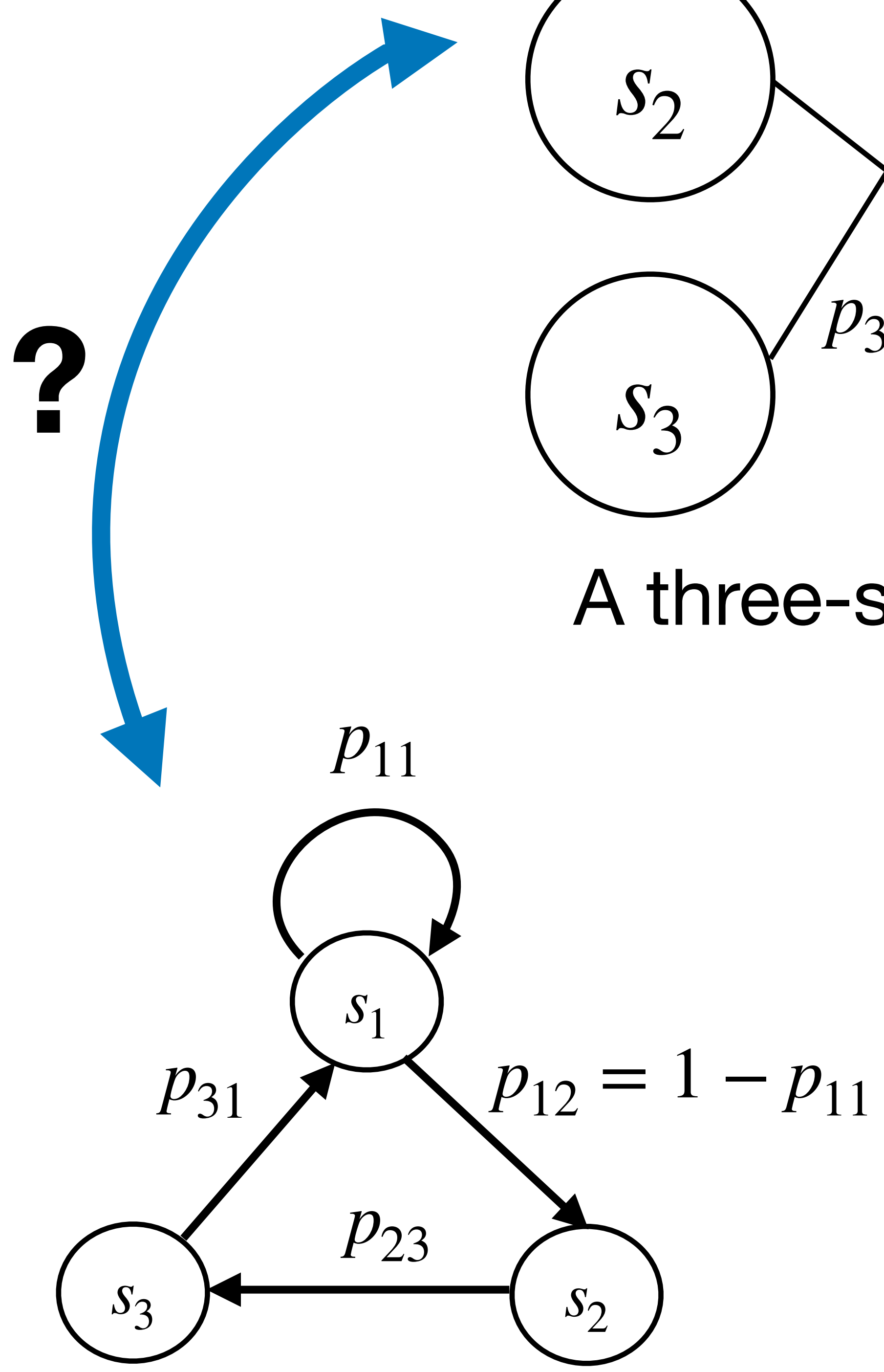
A Markov process vs. a non-Markov process on a countable space



A three-state Markov process transiting from time t to $t+2$



A three-state non-Markov process transiting from time t to $t+2$



Graph representation

- Left: All three states transitions to the next time depend on the current states $\rightarrow$ a Markov process
- Right: The state s_1 and s_2 at time $t+2$ depend on the states at times t and $t+1$ $\rightarrow$ a non-Markov process

Finite Markov chains on countable spaces: Transition probability

Let $\Phi = \{\phi_0, \phi_1, \phi_2, \ldots\}$ be a homogeneous finite Markov chain on a countable space. The set of values (events) from which the Markov states can take realisations is the sigma-algebra $\mathscr{A}$. For any $x_i, x_j \in \mathscr{A}$, $P(x_j, x_i) = \mathrm{Pr}[\phi_t = x_i \,|\, \phi_{t-1} = x_j]$ is called **transition probability**.

- For all $x_i, x_j \in \mathscr{A}$, the transition probability matrix for Markov chain Φ is

$$P_\Phi = \begin{bmatrix} P(x_0, x_0) & P(x_0, x_1) & \ldots & P(x_0, x_i) & \ldots \\ P(x_1, x_0) & P(x_1, x_1) & \ldots & P(x_1, x_i) & \ldots \\ \vdots & \vdots & \ddots & \vdots & \ddots \\ P(x_j, x_0) & P(x_j, x_1) & \ldots & P(x_j, x_i) & \ldots \\ \vdots & \vdots & \ddots & \vdots & \ddots \end{bmatrix}, \ \underbrace{\forall x_i, x_j \in \mathscr{A} : P(x_i, x_j) \geq 0, \ \sum_{x_l \in \mathscr{A}} P(x_i, x_l) = 1}_{\text{Proper transition probability function}}$$

- The n-step transition probability is defined as

$P^n(x_j, x_i) = \mathrm{Pr}[\phi_{t+n} = x_i \,|\, \phi_t = x_j]$

☑ *Recall row-stochastic matrices? P_Φ is also row-stochastic and in general a non-negative matrix!*

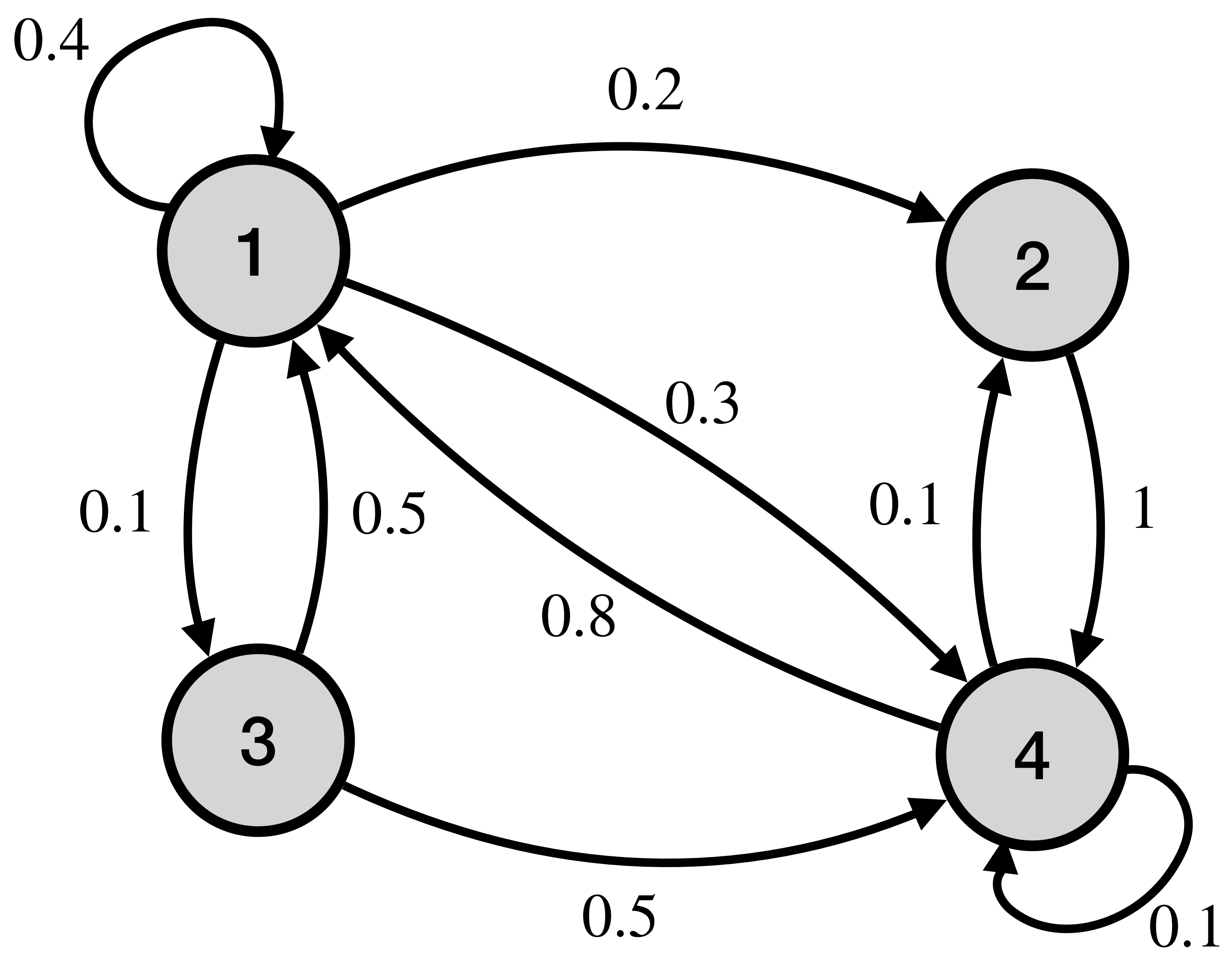
Finite Markov chains on countable spaces: Graph representation

Any time-homogeneous finite Markov chain can be represented by a (row-stochastic) weighted directed graph $G(\mathscr{V}, \mathscr{E}, \{a_e\}_{e\in\mathscr{E}})$, where

- $\mathscr{E} = \{(x_i, x_j)\,|\,P(x_i, x_j) > 0\}$
- $\mathscr{V} = \mathscr{A}$
- $\{a_e\}_{x_i,x_j} = P(x_i, x_j)$

Remember this example? It represents a homogeneous Markov chain Φ where its state takes values from finite set $\mathscr{A} = \{1,2,3,4\}$ with transition probability matrix

$$P_\Phi = \begin{pmatrix} 0.4 & 0.2 & 0.1 & 0.3 \\ 0 & 0 & 0 & 1 \\ 0.5 & 0 & 0 & 0.5 \\ 0.8 & 0.1 & 0 & 0.1 \end{pmatrix}$$



Graph representation of a homogeneous Markov chain Φ

Markov chain Φ equivalently represents

$$x(t+1) = P_\Phi\, x(t)$$

- [ ] Behavior of a dynamical system can be studied with concepts of Markov chains, if it is a Markov process!

Homogeneous Markov chains properties: Irreducibility

- Communication: two states x and y of a Markov chain are said to be **communicating** with each other (commonly shown by $x \leftrightarrow y$) if given transition probability $P(\cdot, \cdot)$

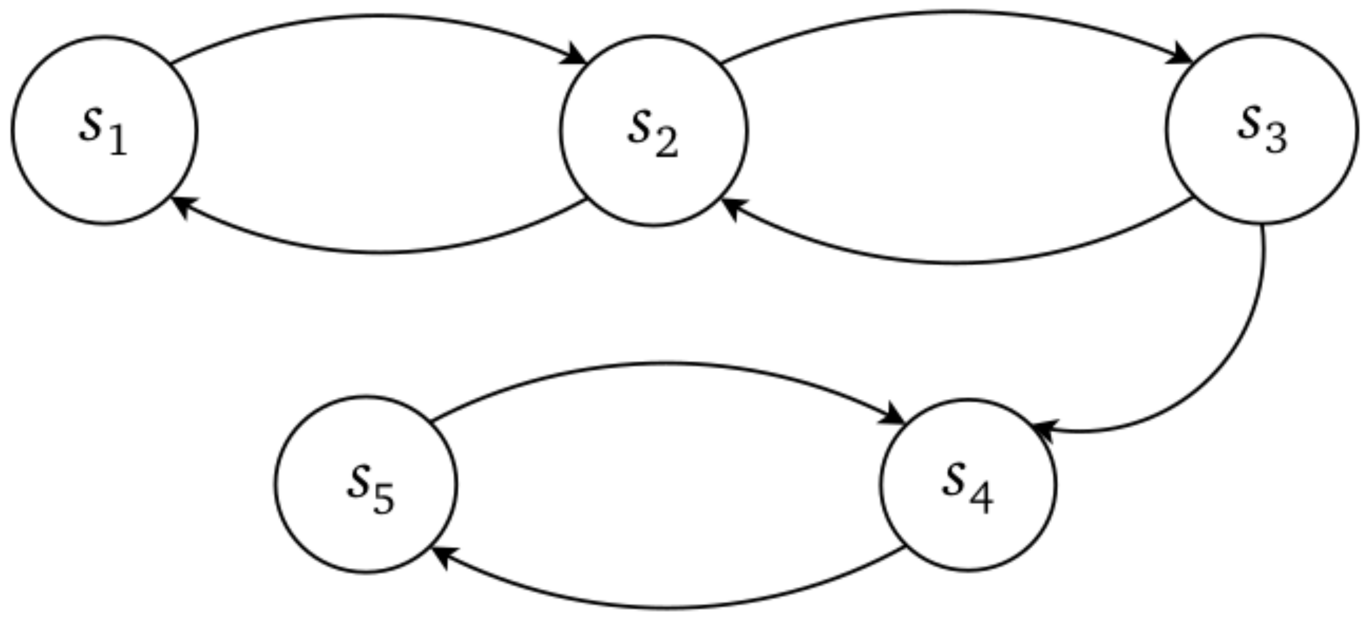
$$\exists n, m \in \mathbb{N}, P^n(x,y) > 0 \;\&\; P^m(y,x) > 0$$

$$\Updownarrow$$

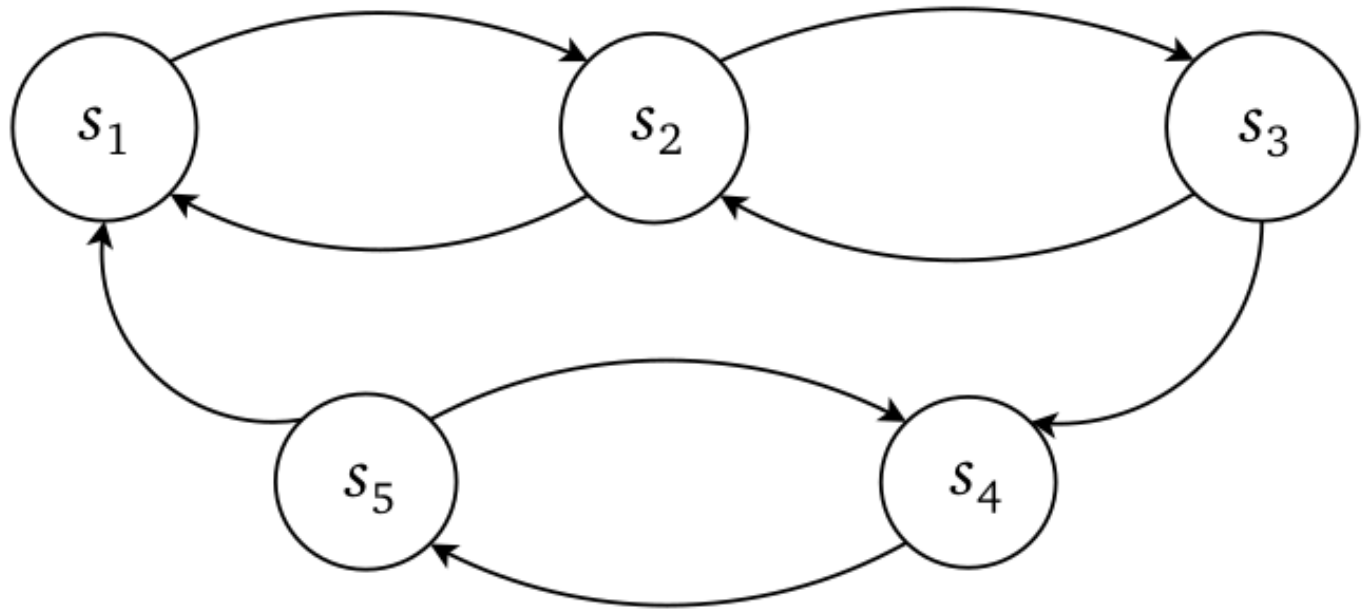
$\exists$ a path of length n (resp. m) from $x \to y$ (resp. $y \to x$) in the graph

- Let a finite Markov chain Φ be defined on a countable space $(\Omega, \mathscr{A}, \mathrm{P})$. If all states of Φ communicate with each other, Φ is said to be **irreducible**

Equivalence: **irreducibility** of a finite Markov chain is equivalent with **strong connectedness** of its representing graph!



A reducible Markov chain on a countable state space



An irreducible Markov chain on a countable state space

Time-homogeneous Markov chains properties: Periodicity

- **Cycle:** Let $\omega \in \Omega$ be a state of a stochastic process defined on Ω. The period of the state ω is defined as

$$d(\omega) = g.c.d.\{n \geq 1 \,|\, p^n(\omega, \omega) > 0\}$$

☑ $p^n(\omega, \omega) > 0 \rightarrow \exists$ a path of length n with starting node ω and ending node ω, i.e., a cycle!

- Let the irreducible Markov chain $\Phi = \{\phi_0, \phi_1, \phi_2, \dots\}$ be defined on a countable space $(\Omega, \mathscr{A}, \mathrm{P})$. Φ is called
  - Aperiodic if $\forall \omega \in \Omega, d(\omega) = 1$, equival. the period of all nodes in the representing graph is one
  - Periodic if $\exists \omega \in \Omega, d(\omega) > 1$, equival. $\exists$ a node in the representing graph with period greater than one
  - Strongly aperiodic if $\exists \omega \in \Omega, P(\omega, \omega) > 0$, equival. $\exists$ a self-cycle in the representing graph

Periodic (left) — with period 2 — vs. aperiodic (right) — with period one — Markov chains (the right chain is also strongly aperiodic since $P(s_1, s_1) > 0$)

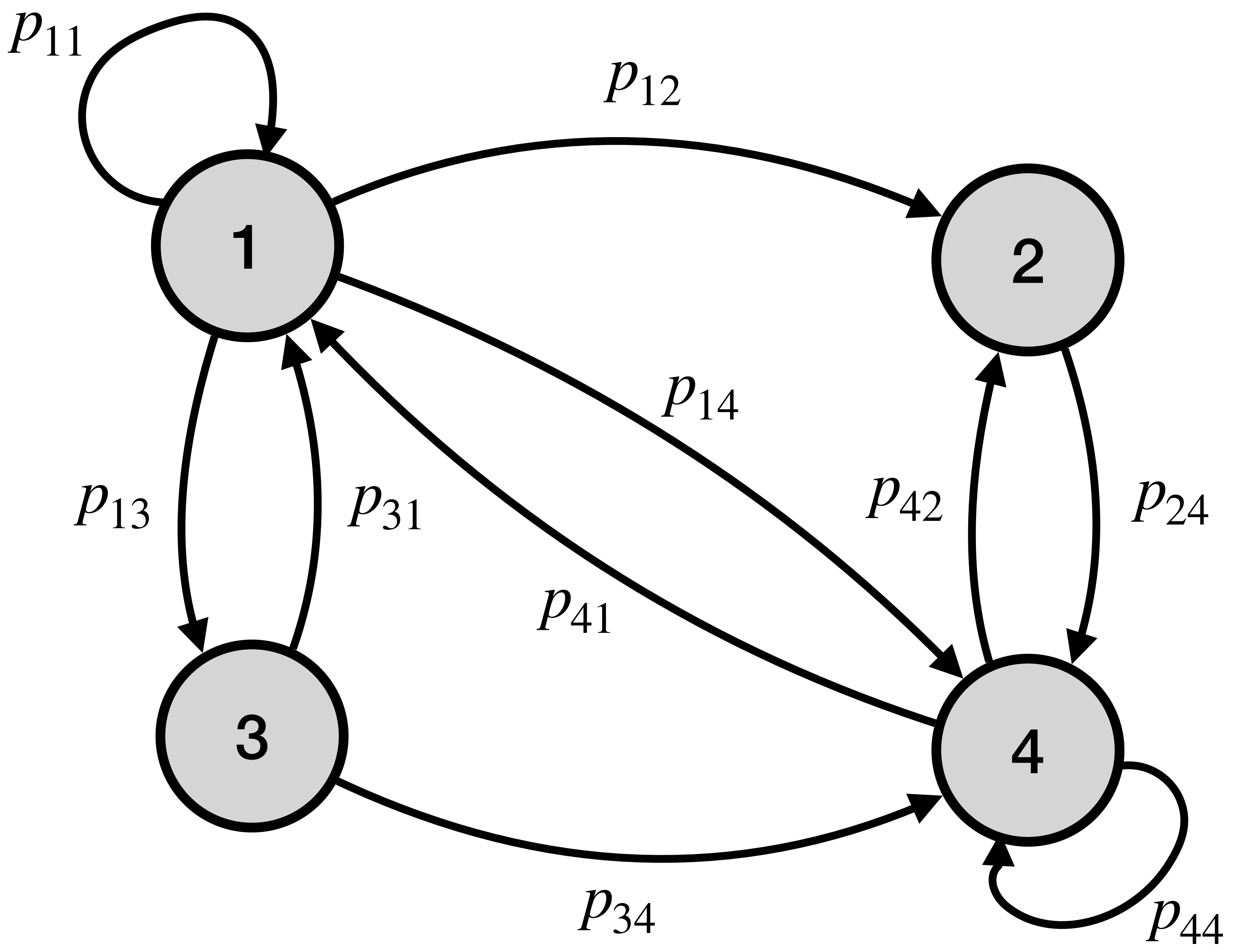
Exercise:

For a finite homogeneous Markov chain with the given graph representation, consider the two following transition probability matrices:

$$P_{\Phi_1} = \begin{pmatrix} 0.5 & 0 & 0.2 & 0.3 \\ 0 & 0 & 0 & 1 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0.4 & 0 & 0.6 \end{pmatrix}, \quad P_{\Phi_2} = \begin{pmatrix} 0 & 0.4 & 0.3 & 0.3 \\ 0 & 0 & 0 & 1 \\ 0.5 & 0 & 0 & 0.5 \\ 0.2 & 0.8 & 0 & 0 \end{pmatrix}$$

Provide answers to the followings:

1. Markov chains reducible or irreducible?
2. Markov chains periodic, aperiodic, or strongly aperiodic?
3. Determine the period of each state.
4. Compute $p^3(2,1)$, $p^2(3,2)$, $p^4(1,4)$ for given transition probability matrices P_{Φ_1} and P_{Φ_2}.



Graph representation of a homogeneous Markov chain Φ

Solution:

Homogeneous Markov chains properties: Occupation time and Return time

Finite irreducible Markov chain $\Phi = \{\phi_0, \phi_1, \phi_2, \ldots\}$ is defined on a countable space $(\Omega, \mathscr{A}, \mathrm{P})$. For any state/event $A \in \mathscr{A}$, we define

- Occupation time η_A: the number of times the chain $\{\phi_1, \phi_2, \ldots\}$ visits A, i.e.

$$\eta_A \triangleq \sum_{n=1}^{\infty} \mathbb{I}\{\phi_n = A\}$$

- First return time τ_A: the first time the chain Φ visits A after time zero, i.e.

$$\tau_A \triangleq \min\{n \geq 1 : \phi_n = A\}$$

For any $A \in \mathscr{A}$:

1. Return time probability $\rightarrow$ $L(A) \triangleq \mathrm{P}(\tau_A < \infty)$
2. Recurrence probability $\rightarrow$ $Q(A) \triangleq \mathrm{P}\{\Phi \in A \text{ infinitely often}\}$, or equivalently, $\eta_A = \infty$

☑ Generally: $Q(A) \leq L(A)$

Stochastic Stability of Markov Processes: Recurrence vs. Transience

Harris recurrent: A state $A \in \mathscr{A}$ is said to be recurrent if the chain visits A infinitely often, i.e.

$$Q(A) = \mathrm{P}(\eta_A = \infty) = 1$$

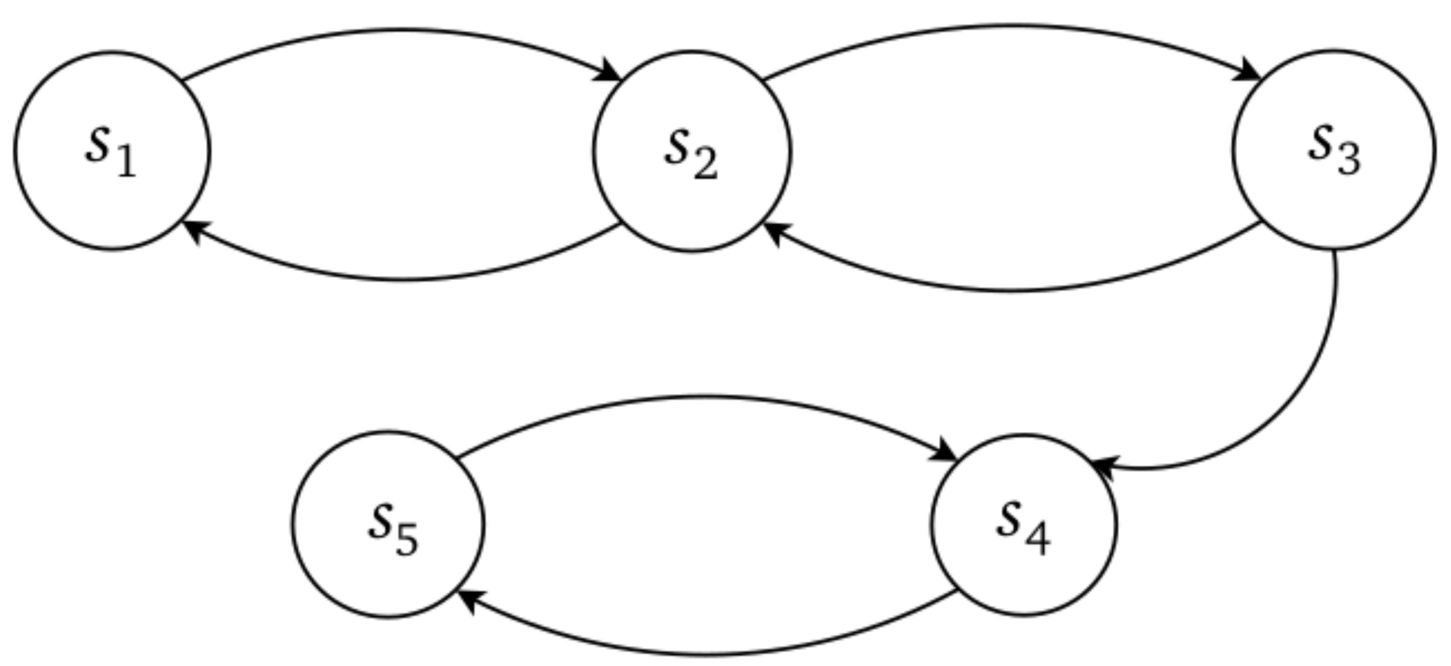
If A is a single state, then we call the state recurrent/transient, if it contains more than one single state, we call it a **recurrent class**!

**If $A \in \mathscr{A}$ is not recurrent, then it is transient!*

Harris recurrent chain: A Markov chain Φ is said to be **positive Harris recurrent** if it is irreducible and

$$L(A) = 1, \forall x \in A, A \in \mathscr{A}$$

i.e., all states in A are Harris recurrent!



A transient discrete-time Markov chain

- Chain return to $\{s_4, s_5\}$ from any other state with non-zero probability! $\rightarrow \{s_4, s_5\}$ is a *recurrent class* and $Q(\{s_4, s_5\}) = 1$
- $\{s_1, s_2, s_3\}$ are transient: if the chain transits to $\{s_4, s_5\}$ it won't return to the *transient class* $\{s_1, s_2, s_3\}$ and $Q(\{s_1, s_2, s_3\}) < 1$

Stochastic Stability of Markov Processes: Recurrence

Theorem: If a Markov chain is defined on a discrete-time finite space, then irreducibility implies recurrence (proof as exercise!)

Irreducibility $\Leftrightarrow L(A) = 1, \forall\, x \in A \in \mathscr{A} \Leftrightarrow Q(A) = 1,\, A \in \mathscr{A}$

Equivalently:

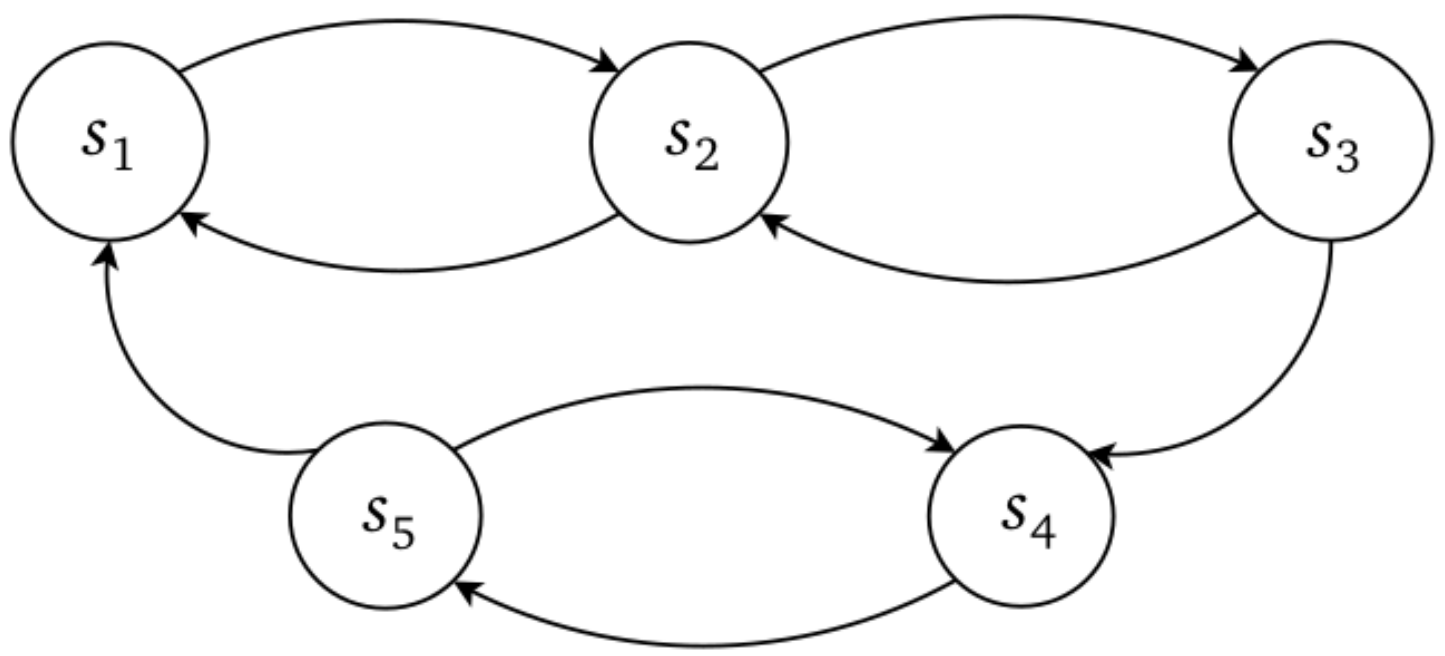
$\exists$ a path from any node to any other node $\Leftrightarrow$ all states reachable from other states with finite transitions $\Leftrightarrow$ all nodes globally reachable in finite time

Example: MC is reducible and no global reachable node (not Harris recurrent!). What are the recurrence probabilities?

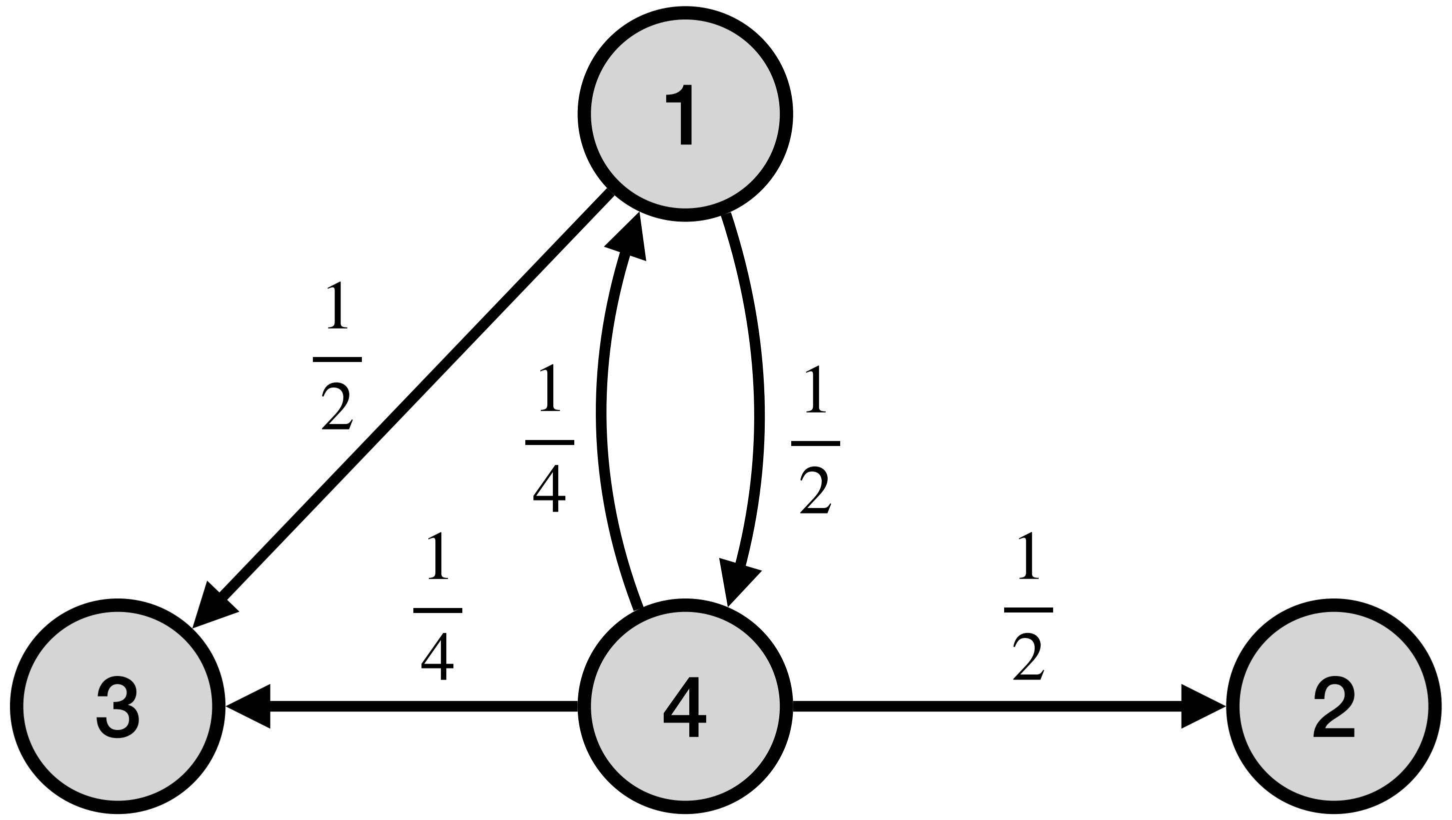
<u>Answer</u>: Recurrent states are 2 and 3:

- $R_2(s_i) =$ P(recurrence to 2 from state i): $R_2(s_2) = 1,\, R_2(s_3) = 0$
- $R_3(s_i) =$ P(recurrence to 3 from state i): $R_3(s_2) = 0,\, R_3(s_3) = 1$

**Calculations on the board!*



A recurrent/irreducible Markov chain on a discrete state space



Graph representation of a homogeneous Markov chain with two recurrent classes

Solution:

Concepts of Stochastic Stability for Markov Chains: Ergodicity

Definition: A homogeneous Markov chain on a countable space is said to be ***ergodic*** if it is irreducible (positive Harris recurrent) and aperiodic

1. If a chain is ergodic: a unique stationary probability distribution Π exists, independent of the initial distributions, such that the Markov chain evolves according to Π when $n \to \infty$, i.e., considering

$$\Pi(A) = \lim_{n\to\infty} \text{Pr}[\phi_n \in A \,|\, \phi_m]$$

If $\Pi(A)$ is finite and independent of $\phi_m, m < \infty$, the chain is ergodic.

☑ If a Markov chain is **periodic** or has **more than one recurrent class**, then it is ***not ergodic***!

2. If a chain is ergodic, the unique probability distribution Π is ***invariant*** $\quad \to \quad \Pi = P_\Phi^n \,\Pi$

Equivalence: Remember consensus algorithm $x(t+1) = Ax(t)$ where A is primitive? $x(\infty) = (w^\top x(0))1_n$

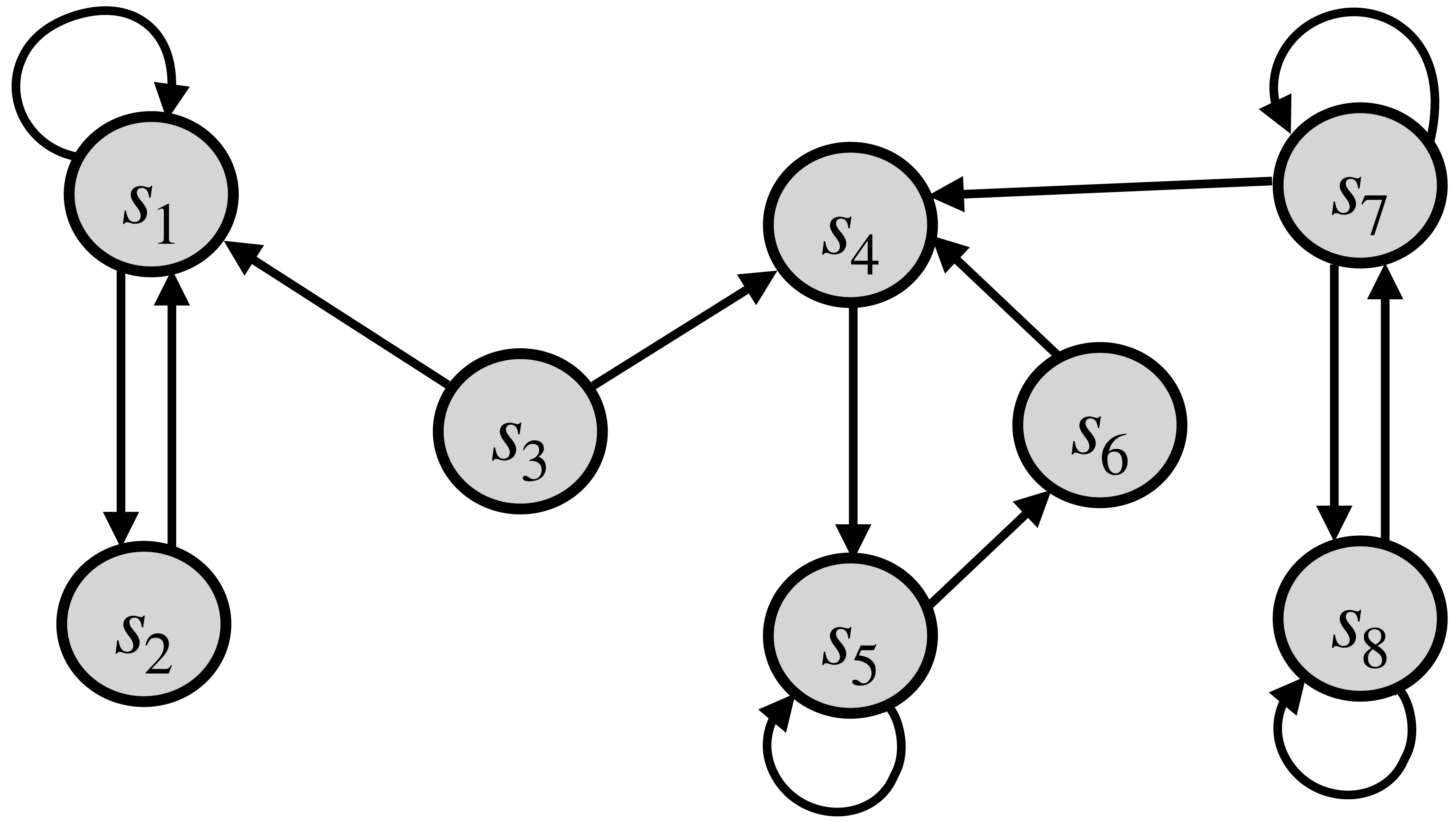
Primitive $\overset{?}{\Leftrightarrow}$ ergodic & $\Pi \overset{?}{\Leftrightarrow} w^\top$

Concepts of Stochastic Stability for Markov Chains: Ergodicity

Example 1: For the given finite state Markov chain we have:

- transient states: $\{s_3, s_7, s_8\}$
- recurrent class 1: $\{s_1, s_2\}$, reachable only from s_3
- recurrent class 2: $\{s_4, s_5, s_6\}$, reachable only from s_3 and s_7

☑ The chain is **not ergodic**! (no globally reachable node)



Intuition: stationary probability of being in a certain state depends on the initial state, e.g., chain starts from s_2, never gets out of the loop $\{s_1, s_2\}$!

☐ Revisit example 5 in Lecture 4 for a graph with no globally reachable node, and with aperiodic SCC sinks (recurrent classes?) What can be concluded?

Example 2: Derive the invariant probability distribution for the given transition probability matrix P_Φ for a 4-state MC

$$P_\Phi = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 \end{pmatrix}$$

** Calculations on the board!*

Stochastic Stability for Markov Chains: Drift criteria

Define the real-valued non-negative measurable function $V(\Phi) : \Omega \mapsto \mathbb{R}^+$. For a discrete state Markov chain $\Phi = \{\phi_0, \phi_1, \ldots, \phi_n, \phi_{n+1}, \ldots\}$ on a countable space, the drift function is defined as:

$$\Delta V(\phi_n) = \mathsf{E}[V(\phi_{n+1}) \,|\, \phi_n] - V(\phi_n), \;\; \phi_n \in \Omega$$

Foster's Criterion: Let the irreducible Markov chain $\Phi = \{\phi_n\}$ be defined on the space Ω with sigma-algebra $\mathscr{A}$. Then Φ is ergodic if a finite set $A \in \mathscr{A}$ exists s.t. for $V(\Phi) : \Omega \mapsto \mathbb{R}^+$, $V(\phi) < \infty, \;\; \phi \in A$:

$$\Delta V(\phi) \leq -1, \;\; \phi \in \Omega \backslash A$$

** Recall the consensus algorithm $x(t+1) = Ax(t)$. If A is primitive $\rightarrow x(t \rightarrow \infty) = \lim_{t\rightarrow\infty} A^t x(0) = (w^\top x(0))1_n$!

Corollary: Let Φ be irreducible and strongly aperiodic. Then Φ is ergodic and $\exists A \in \mathscr{A}$ s.t. for any initial state $x \in \Omega$:

$$\lim_{n\rightarrow\infty} P^n(x, A) - \Pi = 0$$

How about this? $\lim_{n\rightarrow\infty} P^n(x, A) \;\; \Leftrightarrow \;\; \lim_{t\rightarrow\infty} A^t \;\; \rightarrow \;\; \Pi \;\; \Leftrightarrow \;\; 1_n w^\top$