Finite Markov Chains

Vladimir Parfenyev

parfenius@gmail.com

Skolkovo Tech

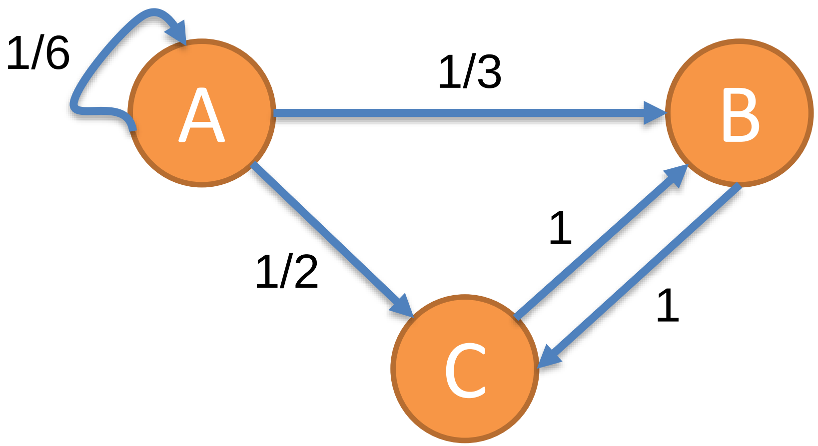
Skolkovo Institute of Science and Technology

14 April 2017

I. Definition & Properties

Finite Markov Chain

Stochastic process with no memory other than of its current state.



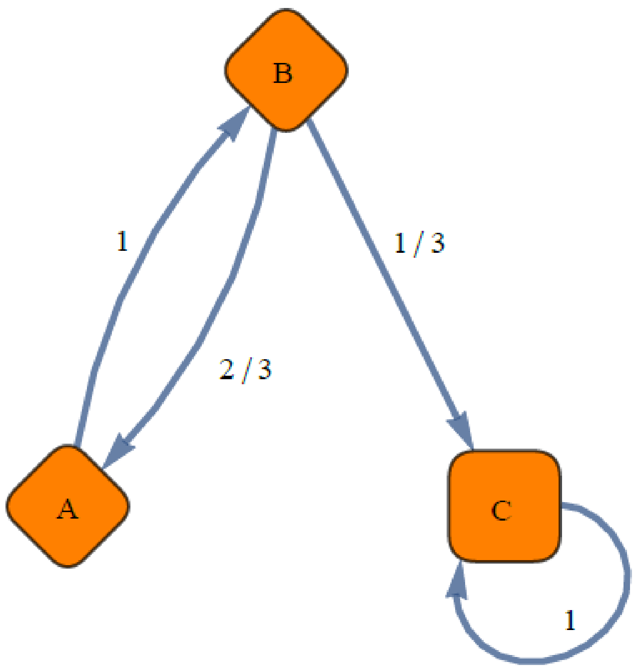
Random walk over a directed graph: vertices – states, edges – transition probabilities $p(i \to j)$.

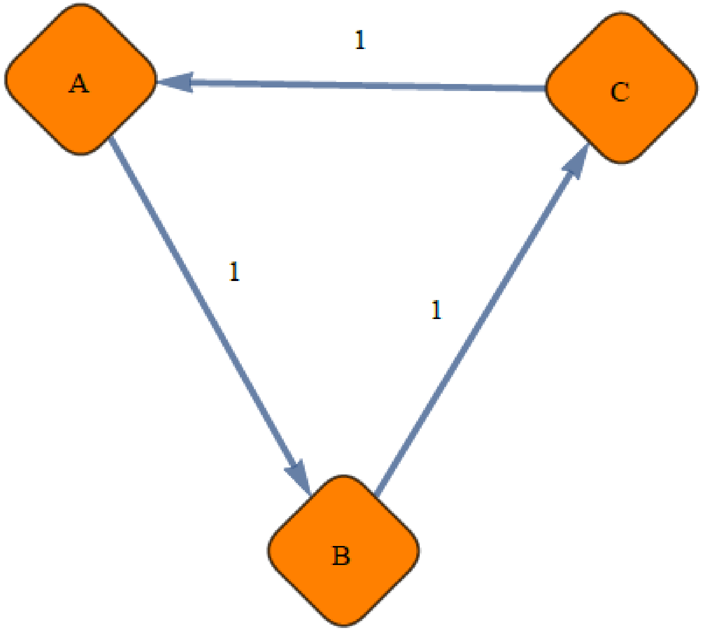
Irreducibility & Reducibility

MC is **irreducible**, if regardless of its present state it reaches, as time progresses, any other state.

Irreducibility & Reducibility

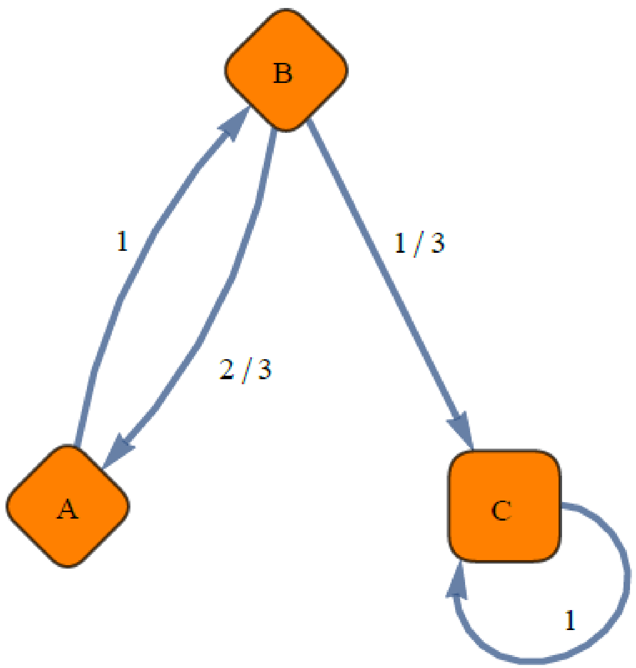
MC is **irreducible**, if regardless of its present state it reaches, as time progresses, any other state.

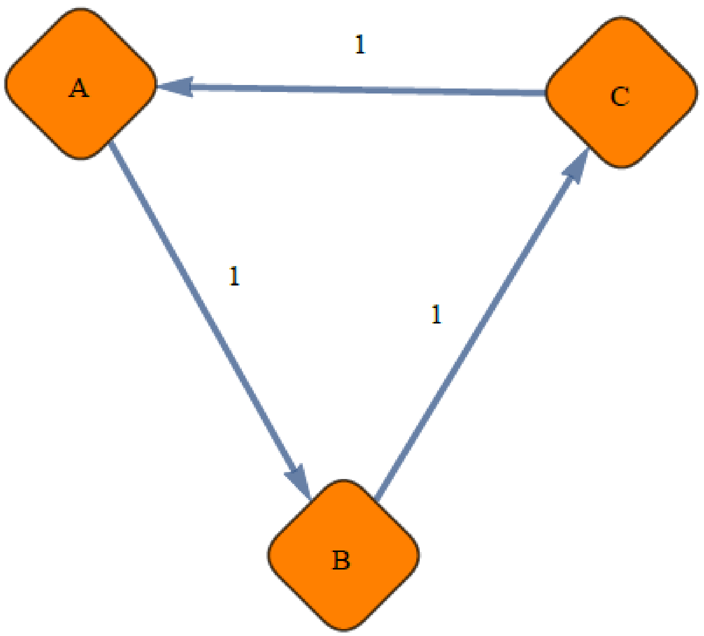




Irreducibility & Reducibility

MC is **irreducible**, if regardless of its present state it reaches, as time progresses, any other state.





Irreducibility is needed to avoid cases with trapped dynamic.

Aperiodicity & Periodicity

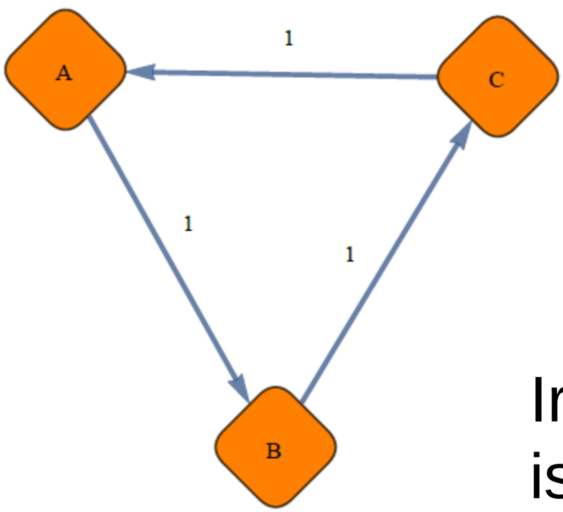
MC is **aperiodic**, if for every state *i* there is *t* such that:

$$\forall t' \geq t \quad Pr(X_{t'} = i | X_0 = i) > 0.$$

Aperiodicity & Periodicity

MC is **aperiodic**, if for every state *i* there is *t* such that:

$$\forall t' \geq t \quad Pr(X_{t'} = i | X_0 = i) > 0.$$



Return to initial state in 3,6,9… steps.
Periodic MC never forgets initial state.

Irreducible MC with at least one **self-loop** is always aperiodic.

Ergodicity

Ergodicity = Irreducibility + Aperiodicity.

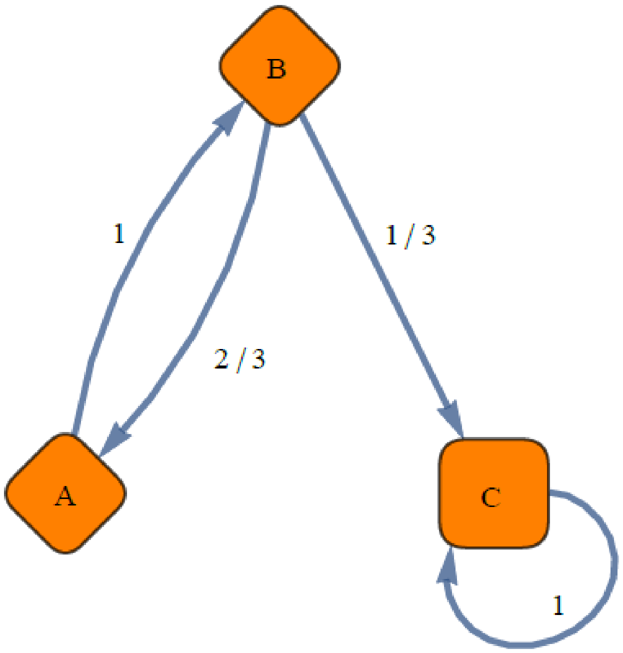
Any ergodic **finite** MC will converge to a **unique** stationary probability distribution, no matter what initial states it starts in.

Ergodicity

Ergodicity = Irreducibility + Aperiodicity.

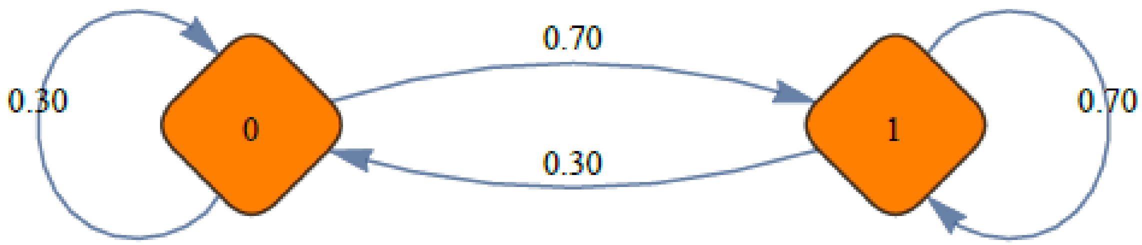
Any ergodic **finite** MC will converge to a **unique** stationary probability distribution, no matter what initial states it starts in.

The opposite statement is not true:



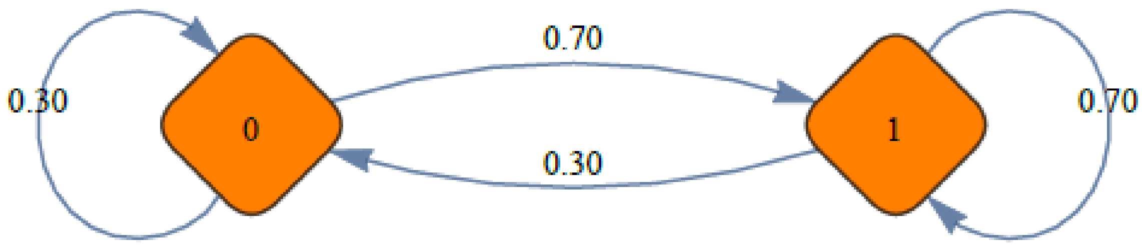
II. Sampling

Idea of Sampling



After some time (for ergodic MC) the probability distribution of a particle becomes stationary (MC is mixed) and the trajectory will represent the sample of distribution.

Idea of Sampling

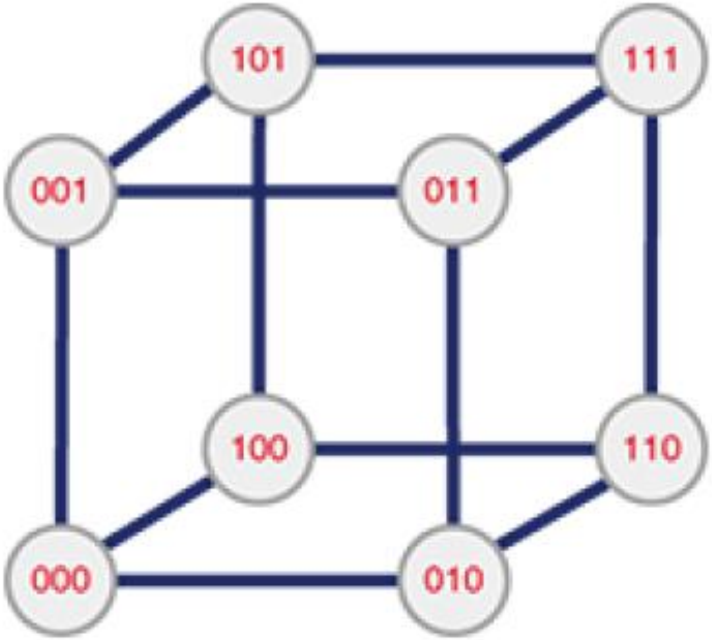


After some time (for ergodic MC) the probability distribution of a particle becomes stationary (MC is mixed) and the trajectory will represent the sample of distribution.

Analyzing the trajectory you can say a lot about distribution, e.g. calculate moments and expectation values of functions.

Walking on Hypercube

Generate random string of *n* bits:

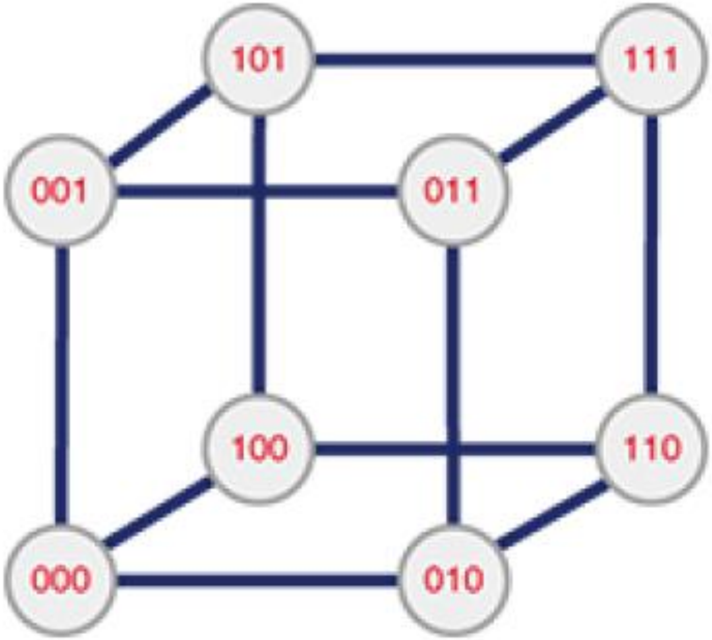


2^n configurations

Lazy Random Walk

Walking on Hypercube

Generate random string of *n* bits:



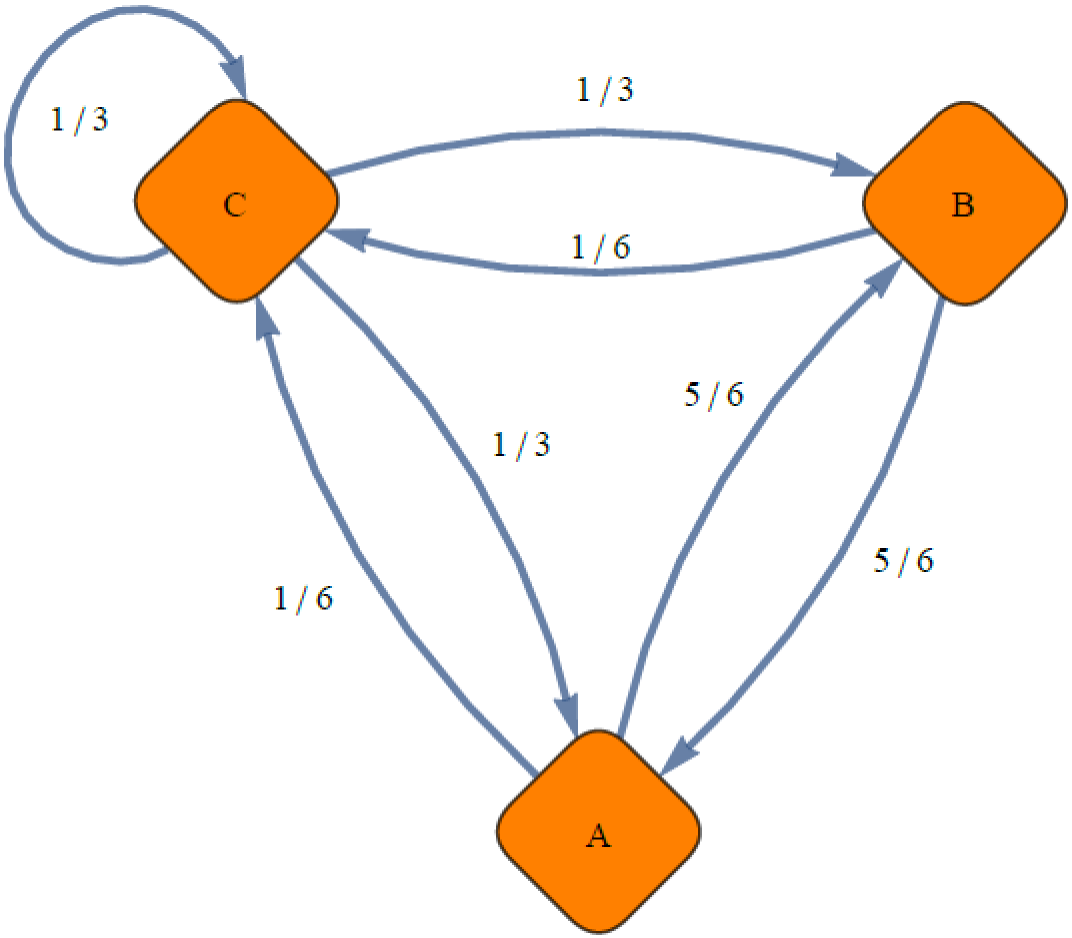
Lazy Random Walk

2^n configurations

How long should we wait before MC becomes mixed (loses memory about initial condition)?

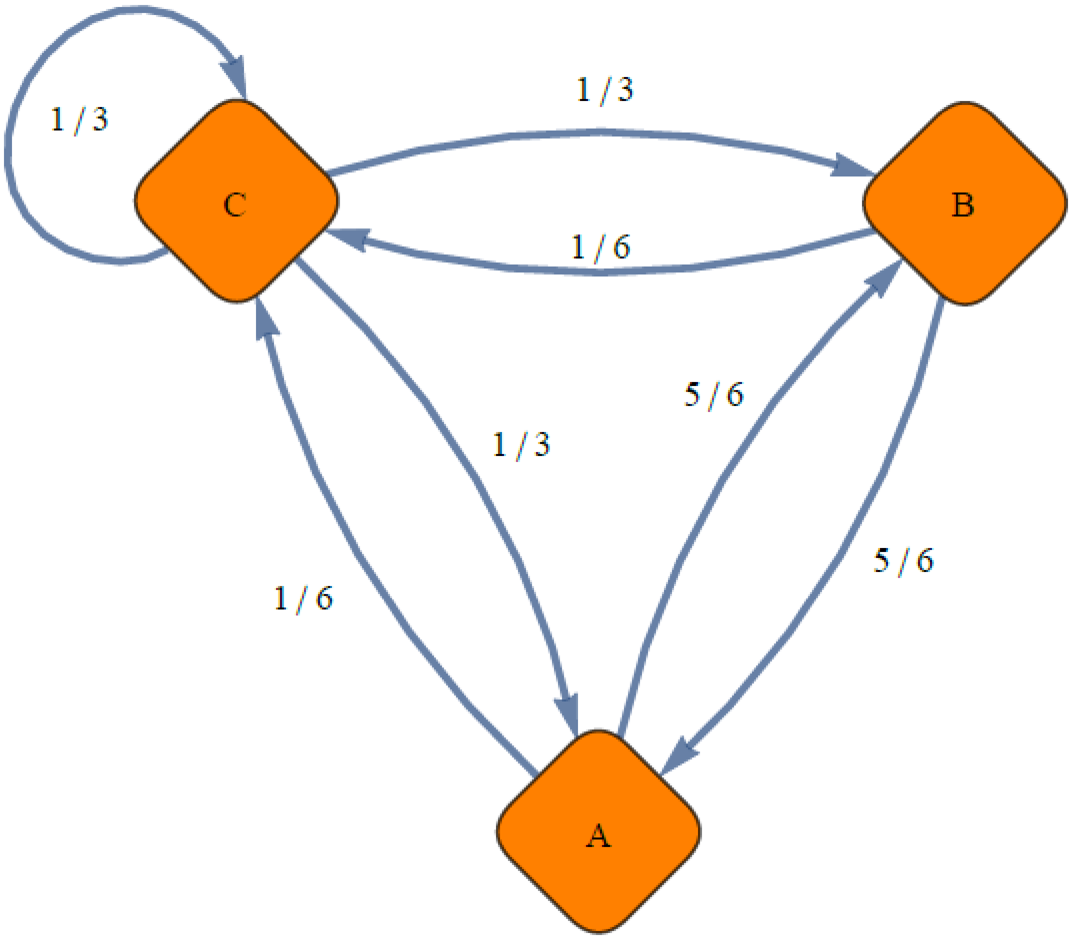
III. Mixing

Matrix Notation



$$\boxed{\pi(t+1) = p\pi(t)}$$

Matrix Notation

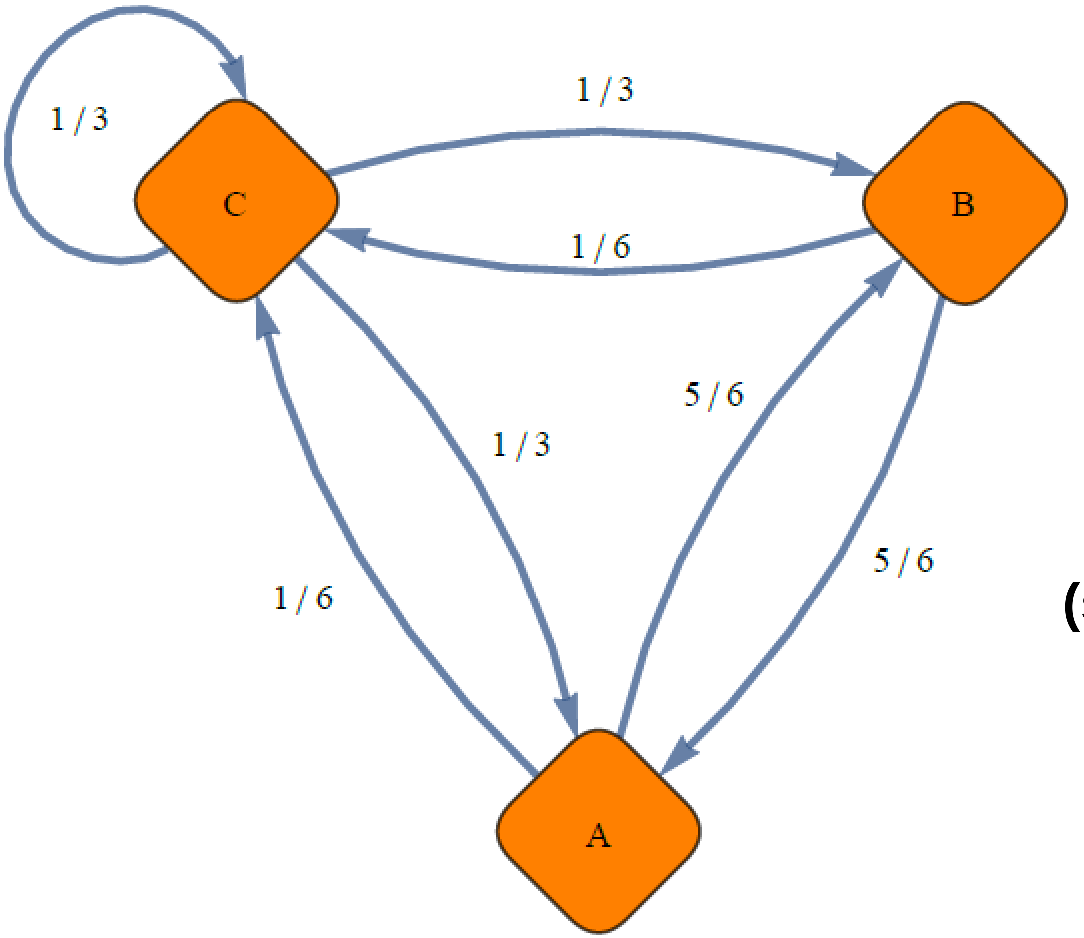


$$\pi_i \geq 0, \ \sum_i \pi_i = 1$$

probability vector

$$\boxed{\pi(t+1) = p\pi(t)}$$

Matrix Notation



$$\pi_i \geq 0, \; \sum_i \pi_i = 1$$

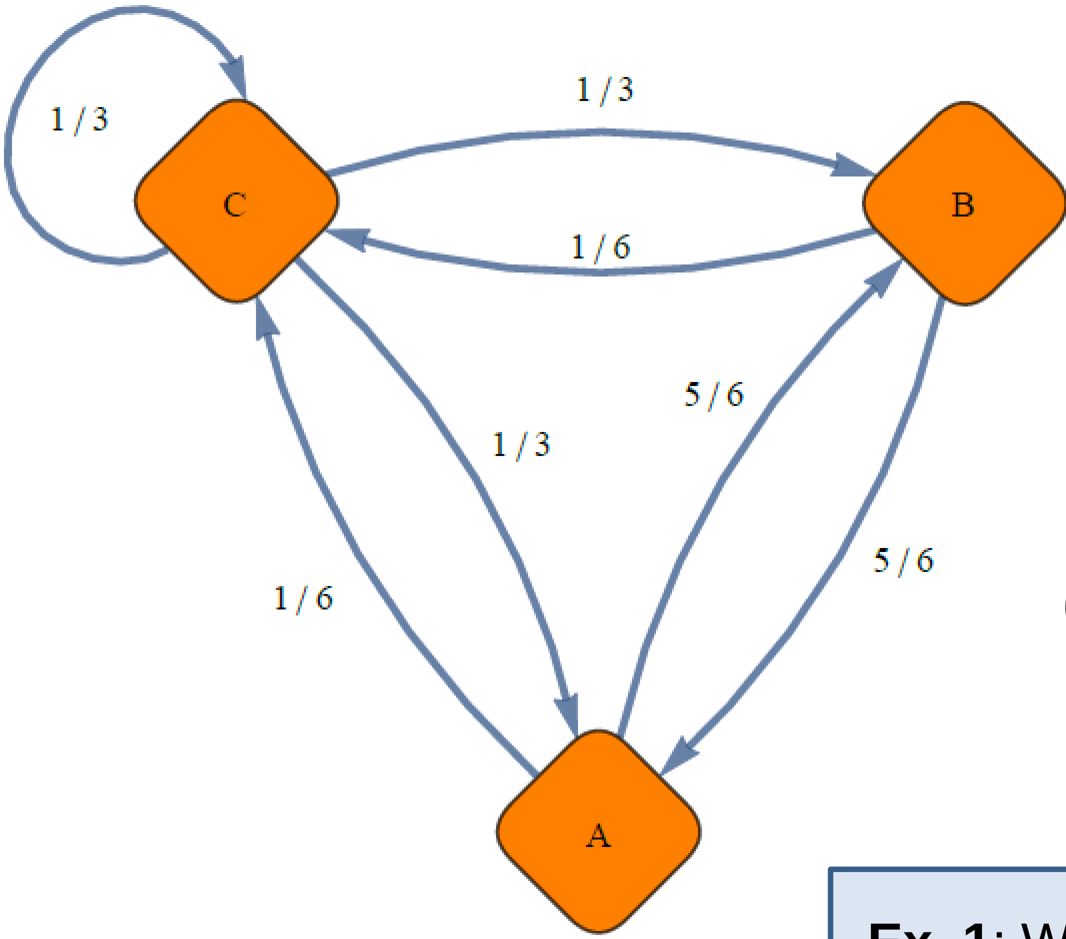
probability vector

$$\boxed{\pi(t+1) = p\pi(t)}$$

(stochastic) transition matrix *p*

$$p_{ij} \geq 0, \; \sum_i p_{ij} = 1$$

Matrix Notation



$$\pi_i \geq 0, \ \sum_i \pi_i = 1$$

probability vector

$$\boxed{\pi(t+1) = p\pi(t)}$$

(stochastic) transition matrix *p*

$$p_{ij} \geq 0, \ \sum_i p_{ij} = 1$$

Ex. 1: Write down transition matrix *p*.

Stationary Distribution

As time increases $\pi(t) = p^t \pi(0)$ approaches stationary distribution

$$p\pi^* = \pi^*.$$

normalized eigenvector $\lambda = 1$

Stationary Distribution

As time increases $\pi(t) = p^t \pi(0)$ approaches stationary distribution

$$p\pi^* = \pi^*.$$

normalized eigenvector $\lambda = 1$

Ex. 2: Find eigenvalues and eigenvectors of *p*. Suppose that we start at the state "A". Write the initial state as a linear combination of eigenvectors and then find the speed of convergence.

Perron-Frobenius Theorem

Ergodic MC with transition matrix *p* has a unique eigenvector with eigenvalue 1, and other eigenvectors have eigenvalues with absolute value < 1.

Perron-Frobenius Theorem

Ergodic MC with transition matrix *p* has a unique eigenvector with eigenvalue 1, and other eigenvectors have eigenvalues with absolute value < 1.

$$1 = |\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \dots$$

defines the speed of convergence

IV. Detailed Balance

Detailed Balance

The distribution satisfies the detailed balance, if for all pairs of states *i, j*

$$\pi_i p(i \to j) = \pi_j p(j \to i).$$

Detailed Balance

The distribution satisfies the detailed balance, if for all pairs of states *i, j*

$$\pi_i p(i \to j) = \pi_j p(j \to i).$$

Ex. 3: Show that if the distribution satisfies DB, then it is a stationary distribution. *Hint: sum DB condition over all states i.*

Detailed Balance

The distribution satisfies the detailed balance, if for all pairs of states *i, j*

$$\pi_i p(i \to j) = \pi_j p(j \to i).$$

Ex. 3: Show that if the distribution satisfies DB, then it is a stationary distribution. *Hint: sum DB condition over all states i.*

Ex. 4: Check that our example satisfies detailed balance.

Balance Condition

The **detailed balance** is not necessary condition for the stationary distribution. The necessary condition is

$$\sum_j (p_{ij}\pi_j - p_{ji}\pi_i) = 0.$$

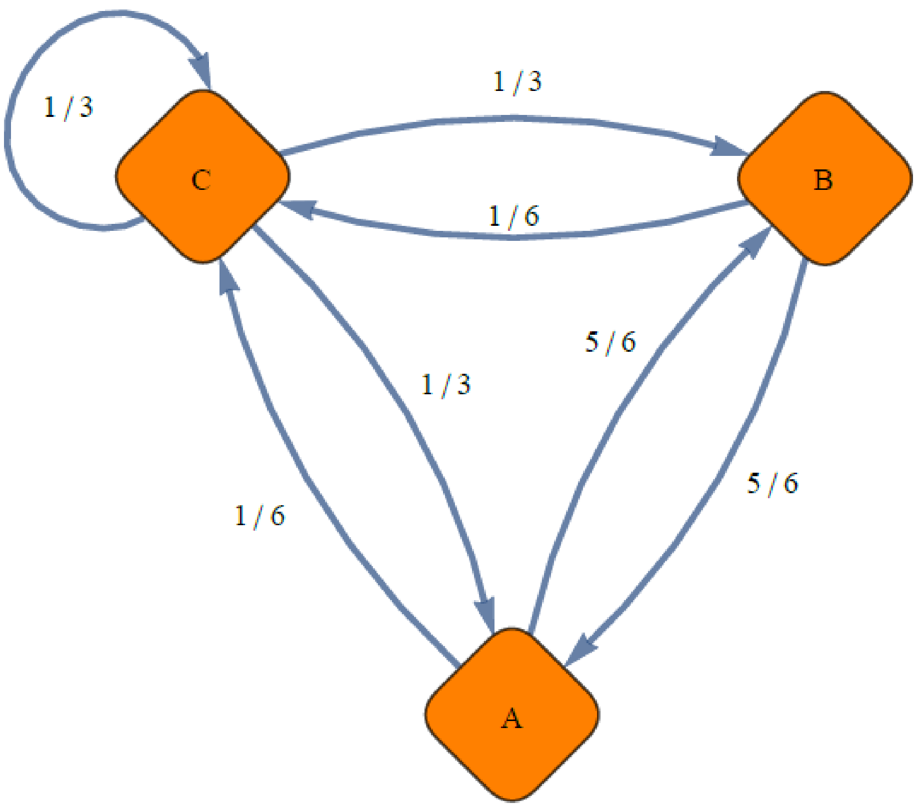
incoming probability flux

outcoming probability flux

V. Efficient Mixing

Problem Formulation

Modify MC to enhance mixing, but we need to preserve the topology of graph and the stationary distribution.



$$p = \begin{pmatrix} 0 & 5/6 & 1/3 \\ 5/6 & 0 & 1/3 \\ 1/6 & 1/6 & 1/3 \end{pmatrix}$$

$$\pi^* = \left(\frac{2}{5}, \frac{2}{5}, \frac{1}{5}\right)^T$$

Hydrodynamic Analogy

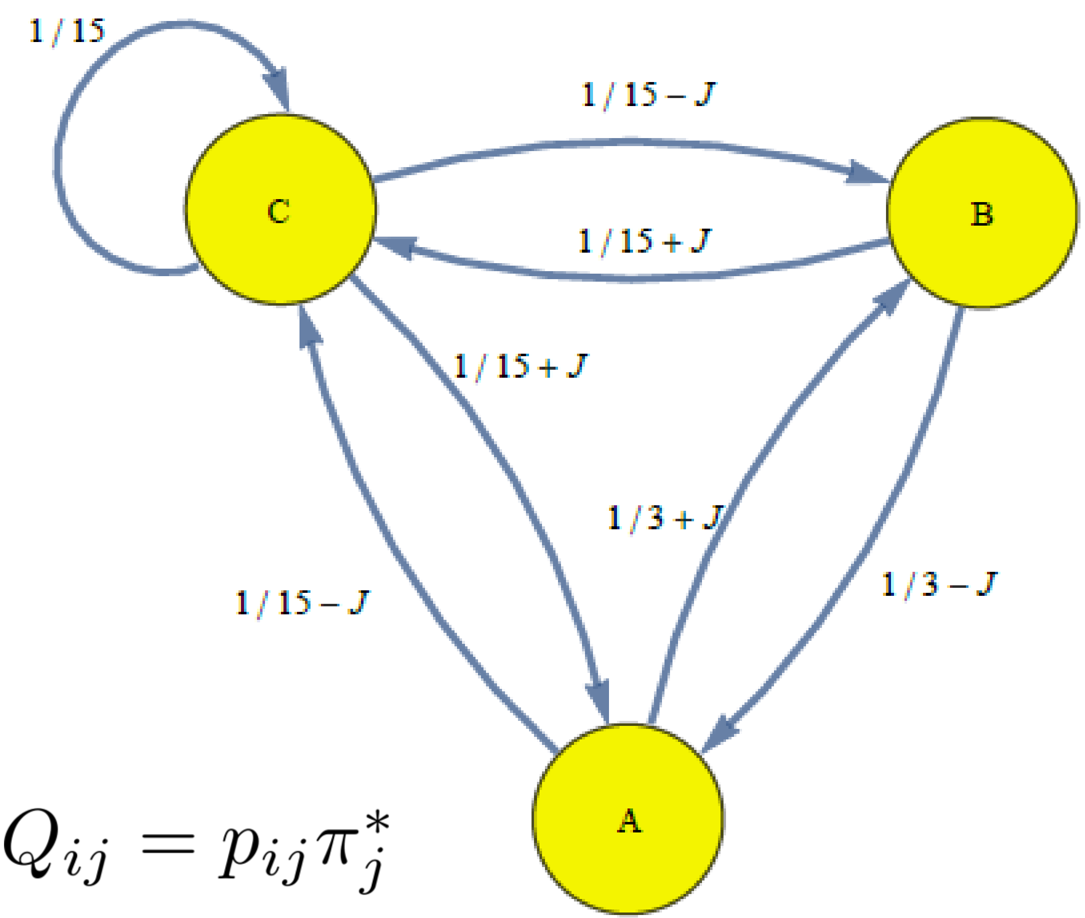
Mixing sugar in a cup of coffee vs Mixing MC

1. Sugar particles have to explore the entire interior of the cup (MC should forget initial condition).

2. Diffusion takes enormous mixing time (analog of detailed balance dynamics).

Better solution – use a spoon!

Probability Fluxes



$$\pi^* = \left(\frac{2}{5}, \frac{2}{5}, \frac{1}{5}\right)^T$$

$$Q_{ij} = p_{ij}\pi_j^*$$

Detailed Balance corresponds to *J* = *0.*

Transition Matrix

$$\tilde{p} = \begin{pmatrix} 0 & 5/6 - 5J/2 & 1/3 + 5J \\ 5/6 + 5J/2 & 0 & 1/3 - 5J \\ 1/6 - 5J/2 & 1/6 + 5J/2 & 1/3 \end{pmatrix} \quad \Rightarrow |J| < 1/15$$

$$\lambda_1 = 1, \quad \lambda_{2,3} = \frac{1}{6}\left(-2 \pm 3\sqrt{1 - 125J^2}\right) \quad \Rightarrow |\lambda_2| = 1/3$$