Available online at www.sciencedirect.com
SciVerse ScienceDirect
Stochastic Processes and their Applications 121 (2011) 2553–2570
stochastic processes and their applications
www.elsevier.com/locate/spa

# Markov chain mixing time on cycles

Balázs Gerencsér*

*Eötvös Loránd University, Pázmány Péter sétány 1/C, Budapest 1117, Hungary*

Received 14 December 2010; received in revised form 14 June 2011; accepted 20 July 2011
Available online 29 July 2011

**Abstract**

Mixing time quantifies the convergence speed of a Markov chain to the stationary distribution. It is an important quantity related to the performance of MCMC sampling. It is known that the mixing time of a reversible chain can be significantly improved by lifting, resulting in an irreversible chain, while changing the topology of the chain. We supplement this result by showing that if the connectivity graph of a Markov chain is a cycle, then there is an $\Omega(n^2)$ lower bound for the mixing time. This is the same order of magnitude that is known for reversible chains on the cycle.
© 2011 Elsevier B.V. All rights reserved.

*Keywords:* Markov chain; Mixing time; Cycle; Non-reversible

## 1. Introduction, result formulation

The goal of this paper is to prove a lower bound on the mixing time of a family of Markov chains.

Mixing time is an important quantity directly related to the performance of numerous algorithms. In Markov chain Monte Carlo simulations (see [8]), mixing time can be interpreted as the time needed to generate a sample.

It turns out that running a local averaging algorithm is the same as following the evolution of the distribution of a certain Markov chain (see [10]). Again, the time needed to get within a certain neighborhood of a common value is quantified by the mixing time.

Motivated by these applications, the estimation of mixing time is in the center of interest.

Usually reversible Markov chains are used to solve these problems. It turns out that often a non-reversible variant can mix much faster. We go a step further on understanding the difference.

* Tel.: +36 208234586.
*E-mail addresses:* gebaboy@cs.elte.hu, balazsg@mit.edu.

0304-4149/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.spa.2011.07.007

Our result is expressed in a single theorem. We work with finite state discrete time Markov chains. We restrict the connectivity graph to a cycle, and allow arbitrary non-reversible transition probabilities such that the uniform distribution is invariant. Then there is a lower bound on the mixing time which has the same order of magnitude as the best lower bound for reversible chains.

We first formulate our result, then we show how it fits into existing literature.

Let us define the quantities and notions we use: If we start the chain with an initial distribution $\sigma$, let $\sigma^{(k)}$ denote the distribution after $k$ steps. For the set of probability distributions on a finite base set $\Omega$ we use the notation $\mathcal{P}(\Omega)$.

**Definition 1.** Given two probability distributions $\mu$ and $\sigma$ on $\Omega$, the total variation distance is

$$\|\mu - \sigma\|_{\mathrm{TV}} = \max_{A \subseteq \Omega} |\mu(A) - \sigma(A)|.$$

**Definition 2.** For a Markov chain with stationary distribution $\pi$ and transition matrix $P = (P_{ij})$, with $P_{ij}$ denoting the probability of moving from state $i$ to state $j$, we define the mixing time of the chain as

$$t_{\mathrm{mix}}(P, \varepsilon) = \max_{\sigma \in \mathcal{P}(\Omega)} \min \left\{ k : \|\sigma^{(k)} - \pi\|_{\mathrm{TV}} \le \varepsilon \right\}.$$

Note that this might be infinite if the Markov chain is non-ergodic.

We consider only the case when the stationary distribution is uniform. For the transition matrix this translates to the condition of being doubly stochastic.

A Markov chain is reversible if starting from the stationary distribution $\pi$, the probability of the consecutive pair $(i, j)$ is the same as the probability of the consecutive pair $(j, i)$. Formally:

$$\pi_i P_{ij} = \pi_j p_{ji} \quad \forall i, j.$$

The connectivity graph of a Markov chain is the graph formed by the states of the Markov chain as nodes and by undirected edges between $i$ and $j$ with $i \neq j$, if either of the transition probabilities $p_{ij}$ or $p_{ji}$ is nonzero. We do not include loops even if $p_{ii} > 0$. We shall also refer to this graph loosely as the topology of the Markov chain. In our case we assume this graph to be a subgraph of a single cycle.

For convenience, let us number the nodes according to the ordering on the cycle. We will interpret these numbers *mod n*.

We are now ready to state our result:

**Theorem 1.** *Consider a Markov chain on a cycle with n nodes having a doubly stochastic transition matrix P. Then, with some global constant $C > 0$ we have*

$$t_{\mathrm{mix}}(P, 1/8) \ge Cn^2.$$

Note that our theorem covers all Markov chains, even non-reversible ones. Quantifiable bounds for the mixing time are often less sharp and/or harder to compute for non-reversible chains. Consider the classic method using the spectral gap $\gamma = (1 - \max_i(|\lambda_i|))$. For lazy reversible chains the mixing time turns out to be roughly $(1/\gamma)\log\varepsilon$, up to a factor of $-\log\min_i \pi_i$. For non-reversible chains, it may happen that the upper bound on TV distance does not converge to 0, so it does not give an upper bound on mixing time at all (see [9]).

The necessity of the separation of reversible and non-reversible Markov chains is widely recognized in literature.

Often it is easier to prove useful properties for reversible chains, and there are tighter general bounds on mixing time for them. The reason to turn to non-reversible chains is the fact that they may deliver much faster mixing than reversible chains with the same connectivity graph. We have to emphasize that this highly depends on the actual connectivity graph. For example if we duplicate a cycle, connect nodes with their pair, then the mixing time can drop to its square root (see [9], Example 6.6). Our theorem aims the other way, it shows that there is no advantage in the case when the topology is a single cycle.

To demonstrate the possibilities, let us cite a result not strictly within our scope, where the connectivity graph actually changes. There is a method to decrease the mixing time of a reversible chain up to its square root by modifying it to a non-reversible one, described in [5,4]. Here the topology of the chain changes as every node is split into multiple copies. Transition probabilities are chosen such that the marginal behaves like the original chain, but we achieve faster mixing on the new graph. The method is called *lifting*.

Although with this method the connectivity graph changes, it is still a powerful example to show what one can achieve. However, the limit of how far we can go is not clear. There are results on finding the fastest mixing reversible chains with fixed connectivity graph, see [3,2], but no such result is available for non-reversible chains.

Our work goes back to the basics. We search for the exact limit of what can be achieved by allowing a non-reversible chain for a given topology, in our case a cycle. It is known that the magnitude of the best mixing time of a reversible chain on a cycle scales with $n^2$ (we will present a proof, see Lemma 9). Our theorem implies that relaxing the reversibility condition does not help with this topology.

The claim of our work is simple to state, however, we did not succeed in proving it using conventional methods. We had to search further and use a unique approach, presented in this paper. As a result, some interesting properties of these Markov chains arise as a by-product.

The rest of the paper is structured in the following way. In Section 2 we prepare the proof and split it into two parts. We have to work on them separately, Sections 3 and 4 deal with these parts. We outline some ideas for future research in Section 5.

## 2. Preparation for the proof

To set up, let us collect some simple observations. First, let us note that in our case of finite state space

$$\|\mu - \sigma\|_{\mathrm{TV}} = \frac{1}{2} \sum_{\omega \in \Omega} |\mu(\omega) - \sigma(\omega)| = \frac{1}{2} \|\mu - \sigma\|_1.$$

We should point it out that the TV distance is defined for measures, $l_1$ is for (real) vectors. In our case we can interpret measures as real vectors, so that this equation makes sense. This means we do not need to use the TV distance but can work with the $l_1$ norm instead. With this change we have to find when the $l_1$ distance decreases below $1/4$ to determine the appropriate mixing time.

Second, let us prove a lemma on the structure of the transition matrix.

**Lemma 1.** *The doubly stochastic transition matrix $P$ of a Markov chain on a cycle can be decomposed as*

$$P = Q + rR,$$

*where $Q$ is the transition matrix of a reversible chain on a cycle, and $R$ is*

$$\begin{pmatrix} & 1 & & & -1 \\ -1 & & 1 & & \\ & -1 & & \ddots & \\ & & \ddots & & \\ & & & & 1 \\ 1 & & & -1 & \end{pmatrix}.$$

**Proof.** Let us start with

$$P = \frac{P + P^T}{2} + \frac{P - P^T}{2} = A + B.$$

The choice $Q = A$ clearly satisfies the conditions we have on $Q$. It is easy to see that $B$ is antisymmetric, and all row and column sums are 0. Set $r = B_{12} = -B_{21}$. Then $B_{23} = -B_{32} = r$ required by the second row and column sum to be zero. Repeat this to get $B = rR$. $\square$

For convenience, we introduce simplified indices for the elements of $Q$ we use often: $q_i = Q_{i-1,i} = Q_{i,i-1}$.

The presence of the $rR$ term has the heuristic effect that the chain is more likely to travel in one direction than the other. This is some sort of rotation, which will play a crucial role in our proof.

Reversing the numbering of the nodes swaps the sign of $r$, so without loss of generality, we may assume $r \geq 0$. Let $\mathcal{M}$ be the set of doubly stochastic transition matrices of a Markov chain on a cycle. Let $\mathcal{M}_0 \subset \mathcal{M}$ be the subset of reversible ones.

Third, let us provide a tool to simplify further discussions:

**Lemma 2.** *Given are a dense subset $\mathcal{N}$ of $\mathcal{M}$, some $\varepsilon > 1/8$ and $K$. Then*

$$\forall P \in \mathcal{N} \quad t_{\mathrm{mix}}(P, \varepsilon) \geq K \Longrightarrow \forall P \in \mathcal{M}\ t_{\mathrm{mix}}(P, 1/8) \geq K.$$

**Proof.** For any matrix $P \in \mathcal{M}$ we have $\|P\|_1 \leq 1$ (the norm is the operator norm w.r.t. the $l_1$ norm). It follows that for any two matrices $P, P' \in \mathcal{M}$,

$$\begin{aligned} \|P^K - P'^K\|_1 &\leq \|P^K - P^{K-1}P'\|_1 + \|P^{K-1}P' - P^{K-2}P'^2\|_1 \\ &\quad + \cdots + \|PP'^{K-1} - P'^K\|_1 \\ &\leq K\|P - P'\|_1. \end{aligned}$$

For any $P \in \mathcal{M}$ choose $P' \in \mathcal{N}$ such that $\|P - P'\|_1 < (\varepsilon - 1/8)/K$. There is an $x \in \mathbb{R}^n$ showing $t_{\mathrm{mix}}(P', \varepsilon) \geq K$. For this $x$,

$$\left\| xP^K - \frac{\mathbf{1}}{n} \right\|_1 \geq \left\| xP'^K - \frac{\mathbf{1}}{n} \right\|_1 - \left\| xP^K - xP'^K \right\|_1 > 2\varepsilon - K\frac{\varepsilon - \frac{1}{8}}{K} > \frac{1}{4}.$$

This confirms $t_{\mathrm{mix}}(P, 1/8) \geq K$. $\square$

We will use this lemma multiple times when we need some extra property for the matrix (such as all eigenvalues are different) which does not hold for all matrices in $\mathcal{M}$. Observe that we can use the lemma independently multiple times if $\mathcal{N}$ is residual each time.

From now on, we have to continue on two tracks. The interesting thing is that we cannot prove Theorem 1 by a single method. In the following two sections we introduce two arguments, one works in the "general" case, when $r > c/n$ and the other works where the chain is almost reversible in the sense that $0 \le r \le c/n$. None of the two arguments can be naturally carried over to the other domain.

The status of $c/n$ is also different in the two parts. In the first part, the value of $c$ is obtained from the proof and is not convenient to change. However, the second argument works for arbitrary $c$. Of course the resulting bound on the mixing time depends on the choice of $c$. Using this flexibility it is enough to prove these two parts as they can be stitched together to cover all possible chains.

## 3. General non-reversible chains

In this section we deal with the case when $r > c/n$, in other words when the chain is "far from reversible".

**Theorem 2.** *Given a Markov chain on an n node cycle consider the doubly stochastic transition matrix $P = Q + rR$ as in Lemma 1. If $r > 2^{11}/n$, then*

$$t_{\mathrm{mix}}(P, 1/8) \ge \frac{1}{2^{12}} n^2.$$

First we give a very short outline of the proof. We use variables not yet defined and relations not yet shown, the point is to sketch the formal structure of the proof.

As a start let us look at a series of vectors $x^l$ approximately following the evolution of the chain:

$$x^l P = x^{l+1} + e^l,$$

with $x^1$ being a probability distribution. Observe that $P$ does not increase the $l_1$ norm, this confirms the following:

$$\begin{aligned}\left\| x^1 P^{k-1} - \frac{\mathbf{1}}{n} \right\|_1 &\ge \left\| x^k - \frac{\mathbf{1}}{n} \right\|_1 - \left\| x^{k-1} P - x^k \right\|_1 \\ &\quad - \left\| x^{k-2} P^2 - x^{k-1} P \right\|_1 - \cdots - \left\| x^1 P^{k-1} - x^2 P^{k-2} \right\|_1 \\ &\ge \left\| x^k - \frac{\mathbf{1}}{n} \right\|_1 - \sum_{l=1}^{k-1} \left\| e^l \right\|_1 . \end{aligned} \tag{1}$$

The left hand side is the quantity we need to keep above $1/4$ as long as possible to ensure a large mixing time. For all $l$ we may use the bound $\|e^l\|_1 < B$, and for an appropriate $k$ we have $\|x^k - \mathbf{1}/n\|_1 > A$. Now using $k \ge n^2/2^{12}$ and $A - kB > 1/4$, we get the bound on the mixing time we are aiming for.

The following things are left.

We have to construct the series $x^l$. It needs to approximately follow the effect of $P$ so that $e^l$ is small. We also want to easily access elements with high indices in order to have a lower bound of the type $\|x^k - \mathbf{1}/n\|_1 > A$. In the end the structure that will give us these vectors will be completely different from a Markov chain, but with the proper tuning it will coincide with it in some sense.

Then we need to prove the lower and upper estimates we used above.

### 3.1. The construction

The main idea is to find $x^l$ in such a way that $x^{l+1}$ is obtained from $x^l$ by a kind of rotation. To define the rotation of a vector we proceed as follows. We consider the unit circle and we fix a function $f$ defined on the circle. We will fix a set of $n$ "observation points" $Z_0, Z_1, \ldots, Z_{n-1}$, and define

$$y_i^0 = f(Z_i).$$

The rotation of the vector $y^0 = (y_i^0)$ is constructed via the rotation of $f$, defined as

$$f^\alpha\left((\cos(u+\alpha), \sin(u+\alpha))\right) = f\left((\cos(u), \sin(u))\right).$$

Then define

$$y^\alpha = (f^\alpha(Z_0), f^\alpha(Z_1), \ldots, f^\alpha(Z_{n-1})).$$

When we use angles we mean them as $mod\ 2\pi$ numbers. Obviously, the vectors $y^\alpha$ need not be probability vectors, so they will have to be normalized. This will be much easier to describe later, let us leave this for now.

Now, let us specify the functions and variables introduced, starting with $f^\alpha$. This is piecewise linear in the angle:

$$f^\alpha\left((\cos(u+\alpha), \sin(u+\alpha))\right) = \left|\frac{u}{2\pi}\right|, \quad u \in [-\pi, \pi).$$

This implies that a rotation by a small angle $\varphi$ would entail a change in $y_i^\alpha$ by an amount of $\pm\varphi/(2\pi)$, except perhaps for the indices corresponding to observation points near $\alpha$ and $\pi + \alpha$.

To achieve a similar effect as this rotation by the Markov dynamics, we need

$$(y^\alpha P)_i - y_i^\alpha = \pm\lambda$$

for some constant $\lambda$, and as many $(\alpha, i)$ pairs as possible. We will not solve this right away, but use it as a motivation. Let us write out the left side:

$$\begin{aligned}(yP)_i - y_i &= y_{i-1}(q_i + r) + y_i(1 - q_i - q_{i+1}) + y_{i+1}(q_{i+1} - r) - y_i \\ &= -(y_i - y_{i-1})(q_i + r) + (y_{i+1} - y_i)(q_{i+1} - r).\end{aligned}$$

Roughly speaking, the use of the functions $f^\alpha$ implies that most $y_i - y_{i-1}$ are proportional to the angular difference of $Z_i$ and $Z_{i-1}$ (neglecting the sign). Let us replace $y_i - y_{i-1}$ with $\delta_i$ in the equation above, and think of $\delta_i$ as this angular difference. We will properly explain this $y - Z - \delta$ relation later.

For the right hand side, let us choose $\lambda = -2r$ (this is a convenient, but arbitrary choice) and drop the sign so that we end up with the system of equations:

$$-\delta_i(q_i + r) + \delta_{i+1}(q_{i+1} - r) = -2r, \quad i = 0, 1, \ldots, n-1. \tag{2}$$

Now the key point is that this system has a positive solution in $\delta_i$. The following lemma ensures this positive solution exists. Once we have $\delta_i$ in our hands, we will properly specify $Z_i$ and thus $y_i$.

**Lemma 3.** *Consider the system of equations*

$$-u_i a_i + u_{i+1} b_{i+1} = -c_i, \quad i = 0, 1, \ldots, n-1.$$

*Suppose $a_i > b_i > 0, c_i > 0$ for $i = 0, 1, \ldots, n-1$. The indices are taken mod $n$. Then the system has a unique, positive solution in $u_i$, $i = 0, 1, \ldots, n-1$.*

**Proof.** We may rearrange the equation to

$$u_{i+1} = u_i \frac{a_i}{b_{i+1}} - \frac{c_i}{b_{i+1}}.$$

This is a linear equation where $u_i$ has a positive coefficient, and a positive constant is subtracted. We can start with $i = 0$ to get an expression for $u_1$ in terms of $u_0$. Then we plug this into $i = 1$, and so on. After going through the full cycle, we end up at

$$u_0 = Au_0 - C.$$

Here $C > 0$ because it is the sum of positive numbers, and $A = \frac{\prod_{i=0}^{n-1} a_i}{\prod_{i=0}^{n-1} b_i} > 1$. So the solution $u_0 = C/(A-1)$ is positive. Plugging this back allows us to compute all other $u_i$ and we just made sure that it will be consistent when we arrive back to $u_0$.

Suppose $u_i \leq 0$ for some $i$. From the equation it follows that $u_{i+1} < 0$. If we continue this we find $u_0 < 0$ which is impossible, so indeed $u_i > 0$.

Uniqueness is clear by the method we described. $\square$

Let us add up Eq. (2) for all $i$. A lot of terms cancel out and we get

$$\sum_{i=0}^{n-1} \delta_i = n.$$

As we said before, we want these $\delta_i$ to be proportional to the angles between $Z_i$'s. In order to fit these on the circle, we have to scale them down. Let $Z_0$ be the point at angle 0, and $Z_i$ be the point at angle $2\pi \sum_{j=1}^{i} \delta_j / n$.

Let us check the construction. On the half circle where $f$ increases with the angle $y_i^\alpha - y_{i-1}^\alpha = \delta_i / n$, by Eq. (2),

$$(y^\alpha P)_i - y_i^\alpha = -2r/n.$$

The same happens on the other half but with opposite signs. Naturally the nodes near $\alpha$ and $\pi + \alpha$ may behave differently, and we have to make sure they stay under control. This change of $\pm 2r/n$ corresponds to a $4\pi r/n$ angle rotation of $f^\alpha$. This justifies the definition of the error term

$$d^\alpha = y^\alpha P - y^{\alpha + \frac{4\pi r}{n}}. \tag{3}$$

After describing our variables we need to prove the bounds used in the outline of the proof.

### 3.2. Bounds on errors

First we prove a bound on $\delta_i$.

**Lemma 4.** *For every $i$,*

$$\delta_i \leq \frac{2}{q_i}.$$

**Proof.** Let us start from Eq. (2) on $\delta_i$. We can write it in the following way:

$$(q_i + r)\delta_i - (q_{i+1} - r)\delta_{i+1} = 2r.$$

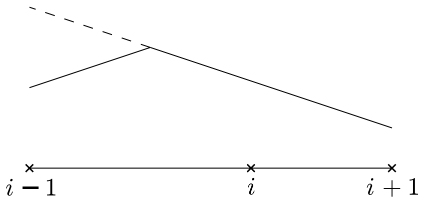


Fig. 1. Node near the peak.

If $\delta_i > 2$ (or $\delta_{i+1} > 2$) it follows that

$$q_i\delta_i - q_{i+1}\delta_{i+1} < 0.$$

Now suppose $\delta_i > \frac{2}{q_i}$. This is clearly more than 2, so we have

$$\delta_{i+1} > \frac{q_i}{q_{i+1}}\delta_i > \frac{2}{q_{i+1}}.$$

We can continue this argument for the next index:

$$\delta_{i+2} > \frac{q_{i+1}}{q_{i+2}}\delta_{i+1} > \frac{q_{i+1}}{q_{i+2}}\frac{q_i}{q_{i+1}}\delta_i = \frac{q_i}{q_{i+2}}\delta_i > \frac{2}{q_{i+2}}.$$

After doing this $n$ times, we end up with $\delta_i > \delta_i$ which is a contradiction, so the claim of the lemma is indeed true. □

The previous lemma helps to bound $d^\alpha$. This $d^\alpha$ will become the error term $e^l$ used in the outline of the proof after proper scaling.

**Lemma 5.** *For every $\alpha$,*

$$\|d^\alpha\|_1 \le \frac{24}{n}.$$

**Proof.** If we pick a node $i$, and $f^\alpha$ is linear on the joint arc between $Z_{i-1}$ and $Z_{i+1}$, things work as we designed them, and $(y^\alpha P)_i - y_i^{\alpha+\frac{4\pi r}{n}} = 0$. There are two irregular arcs, those containing $\alpha$ and $\pi+\alpha$, this affects at most four nodes. Let us focus on these nodes (see Fig. 1).

There would be no error at node $i$ if we used the dashed line, so we have to measure the difference caused by switching to the real, solid line.

The slope of the line is $1/(2\pi)$ so the difference at $y_{i-1}^\alpha$ is at most $\delta_i/n$. During the rotation, the peak might reach $Z_i$ so the value of $y_i^{\alpha+\frac{4\pi r}{n}}$ might deviate at most $4r/n$ from the dashed line. Adding up these two sources of error we get

$$\left|(y^\alpha P)_i - y_i^{\alpha+\frac{4\pi r}{n}}\right| \le \frac{\delta_i}{n}(q_i + r) + \frac{4r}{n}.$$

Let us note $q_i + r$ and $q_i - r$ are both transition probabilities, thus $r \le q_i$ and $r \le 1/2$.

$$\left|(y^\alpha P)_i - y_i^{\alpha+\frac{4\pi r}{n}}\right| \le 2\frac{\delta_i}{n}q_i + \frac{2}{n} \le \frac{6}{n}.$$

The last inequality follows from Lemma 4. The same bound is true if the peak is between $Z_i$ and $Z_{i+1}$. Adding four of these and a few zeros proves the lemma. □

*3.3. Bounds on initial distance*

Although we want to use $y^\alpha$ for $x^1$, it is generally not a probability distribution, so we have to figure out how to scale it. Observe that

$$y^\alpha + y^{\pi+\alpha} = \frac{\mathbf{1}}{2}.$$

Consequently $\|y^\alpha\|_1 + \|y^{\pi+\alpha}\|_1 = \frac{n}{2}$. The value $\|y^\beta\|_1$ is continuous in $\beta$, so we can choose $\beta$ such that

$$\left\|y^\beta\right\|_1 = \frac{n}{4}. \tag{4}$$

This $\beta$ will be fixed from now on, and also $x^1 = \frac{4}{n}y^\beta$, which is now a valid probability distribution.

The last building block of the proof can be summarized in the following lemma:

**Lemma 6.** *Suppose the assumptions of Theorem 2 hold. Then there exists a $k \in [\frac{1}{2^{12}}n^2 + 1, \frac{1}{2^{11}}n^2 + 1]$ such that*

$$\left\|\frac{4}{n}y^{\beta+k\frac{4\pi r}{n}} - \frac{\mathbf{1}}{n}\right\|_1 > \frac{1}{3}.$$

We need some simple lemmas to prove this. Let us introduce the notation

$$s(\alpha) = \left\|\frac{4}{n}y^\alpha - \frac{\mathbf{1}}{n}\right\|_1.$$

**Lemma 7.** *The function $s$ cannot change too fast:*

$$|\overline{s}'(\alpha)| \le \frac{2}{\pi}, \quad |\underline{s}'(\alpha)| \le \frac{2}{\pi} \quad \forall \alpha \in [0, 2\pi).$$

*Here $\overline{s}'$ and $\underline{s}'$ are the upper and lower derivatives, respectively. On the other hand, for the average value:*

$$\frac{1}{2\pi}\int_0^{2\pi} s(\alpha)\mathrm{d}\alpha = \frac{1}{2}.$$

**Proof.** The derivative of $\frac{4}{n}y_i^\alpha$ is in $[-\frac{2}{n\pi}, \frac{2}{n\pi}]$. This also holds if we subtract a constant and take the absolute value. If we add up $n$ of these, we get exactly what we stated.

For the second claim,

$$\frac{1}{2\pi}\int_0^{2\pi}\left|\frac{4}{n}y_i^\alpha - \frac{1}{n}\right|\mathrm{d}\alpha = \frac{1}{2}\int_0^2\left|\frac{u}{n} - \frac{1}{n}\right|\mathrm{d}u = \frac{1}{2n}.$$

Adding these up gives the second formula. □

**Lemma 8.** *The function $s(\alpha)$ is continuous and piecewise linear with at most $4n$ segments on $[0, 2\pi]$, assuming 0 and $2\pi$ are stitched together.*

**Proof.** Again,

$$s_i(\alpha) = \left|\frac{4}{n}y_i^\alpha - \frac{1}{n}\right|$$

is piecewise linear with four segments (four points of nonlinearity). If we add up $n$ of such functions, it will have at most $4n$ segments. □

Now we turn back to the lemma we left over.

**Proof of Lemma 6.** Suppose the claim does not hold. Let us mark the set $G$ of "good" points in the following sense:

$$G = \left\{\alpha \in [0, 2\pi], s(\alpha) > \frac{1}{3}\right\}.$$

Let us look at Lemma 8. While we go around the circle on each segment we might step in or out of $G$, but at most once. This means $G$ is the union of at most $2n$ intervals.

On the range of $k$ we are working on, we are rotating

$$\frac{n^2}{2^{12}}\frac{4\pi r}{n} = 2\pi\frac{rn}{2^{11}} > 2\pi.$$

In other words, we rotate through the whole circle. This is the point where we use the lower bound on $r$. If the claim does not hold, it means we never hit $G$ as $k$ sweeps its range. This means we jump over every interval when we reach it. Consequently each interval is at most $\frac{4\pi r}{n}$ long.

At both ends of such an interval $s(\alpha) = 1/3$, so by the bound in Lemma 7 $s(\alpha)$ can increase up to at most $\frac{1}{3} + \frac{2\pi r}{n}\frac{2}{\pi}$ on such a short interval. We can construct an upper estimate on the average distance using different bounds on $G$ and outside $G$. Using Lemma 7 again for the average value gives us the following:

$$2\pi \cdot \frac{1}{2} \le \frac{4\pi r}{n}2n \cdot \left(\frac{1}{3} + \frac{2\pi r}{n}\frac{2}{\pi}\right) + \left(2\pi - \frac{4\pi r}{n}2n\right) \cdot \frac{1}{3}.$$

Rearranging this gives

$$r \ge \sqrt{\frac{n}{96}}.$$

We know $r$ is at most $1/2$. By the condition $r > 2^{11}/n$ we also have $n > 2^{12}$. But then the right hand side becomes more than $1/2$ and this leads us to a contradiction. □

### 3.4. *The proof*

It only remains to put things together. Based on Eq. (4) we defined $x^1 = \frac{4}{n}y^\beta$ as the starting probability distribution. The scaled versions of the rotated vectors are

$$x^l = \frac{4}{n}y^{\beta+(l-1)\frac{4\pi r}{n}}.$$

The error terms bounded in Lemma 5 scale by the same factor, thus we may define

$$e^l = \frac{4}{n}d^{\beta+(l-1)\frac{4\pi r}{n}}.$$

Using these notations, Eq. (3) defining $d^\alpha$ becomes

$$e^l = x^l P - x^{l+1}.$$

Recall we started from Eq. (1):

$$\left\| x^1 P^{k-1} - \frac{\mathbf{1}}{n} \right\|_1 \geq \left\| x^k - \frac{\mathbf{1}}{n} \right\|_1 - \sum_{l=1}^{k-1} \|e^l\|_1.$$

Let us choose $k$ from Lemma 6, then the first term on the right hand side is more than $1/3$. By the definition of $e^l$ and Lemma 5 we have $\|e^l\|_1 < 96/n^2$. Plugging these in and using the bound on $k$ gives

$$\left\| x^1 P^{k-1} - \frac{\mathbf{1}}{n} \right\|_1 > \frac{1}{3} - (k-1)\frac{96}{n^2} \geq \frac{1}{3} - \frac{3}{64} > \frac{1}{4}.$$

Consequently

$$t_{\text{mix}}(P, 1/8) \geq k - 1 \geq \frac{1}{2^{12}} n^2.$$

## 4. Almost reversible chains

In this section we will prove the following:

**Theorem 3.** *Given a Markov chain on an n node cycle consider the doubly stochastic transition matrix $P = Q + rR$. Suppose $0 \leq r \leq c/n$ for some fixed $c > 0$. Then there is a $c' > 0$ such that*

$$t_{\text{mix}}(P, 1/8) \geq c' n^2,$$

*and $c'$ depends only on $c$.*

The idea is to compare our chain to a reversible one. We try to estimate the errors when $r$ is small enough. We do this first with an additional condition on the chain, but we will be able to relax it later.

### 4.1. The reversible case

Let us see how does the proof go if the transition matrix is symmetric. Our argument will be slightly different and more constructive than the usual eigenvalue estimation.

To reduce complexity, we state and prove Lemma 9 only if $n$ is even. The same argument works for the odd case, we only have to do trivial adjustments.

**Lemma 9.** *Suppose $Q$ is as before, $n$ is even. Then for the initial distribution*

$$x = \frac{4}{n^2}\left(0, 1, \ldots, \frac{n}{2} - 1, \frac{n}{2}, \frac{n}{2} - 1, \ldots, 2, 1\right),$$

*some global $c_1 > 0$ and for any $k \leq c_1 n^2$ we have the bound*

$$\left\| x Q^k - \frac{\mathbf{1}}{n} \right\|_1 > \frac{5}{12}.$$

*Consequently,*

$$t_{\text{mix}}(Q, 1/8) > c_1 n^2.$$

**Proof.** Let us consider the vector

$$x_0 = \frac{4}{n^2}\left(0, 1, \ldots, \frac{n}{2} - 1, \frac{n}{2}, \frac{n}{2} - 1, \ldots, 2, 1\right) - \frac{\mathbf{1}}{n}.$$

This is almost the same as $x$, where $4/n^2$ is chosen to normalize the vector in parentheses to a probability distribution. Then we subtract the uniform distribution to make $x_0$ orthogonal to it. (If $n$ was odd, the maximal coordinate would be $(n+1)/2$ and we would have an extra 0 in the end.)

We will split $x_0 Q^k$ into two components. One pointing in the $x_0$ direction, providing the vector is far from uniform, and another perturbing this. We want the first to be large, the second to be small. Let us start estimating the first.

It is well known that the Laplacian of the chain is $I - Q$ and that

$$x_0(I - Q)x_0^T = \frac{1}{2}\sum_{i,j}(x_{0,i} - x_{0,j})^2 Q_{ij}.$$

The nonzero terms of this sum are $16/n^4 \cdot Q_{ij}$. If we add these up, we get

$$x_0(I - Q)x_0^T = \frac{8}{n^3}.$$

On the other hand $x_0 x_0^T = 1/(3n) + 8/(3n^2) > 1/(3n)$, so it follows that

$$\frac{x_0(I - Q)x_0^T}{x_0 x_0^T} < \frac{24}{n^2}.$$

Using Lemma 2 we may assume all eigenvalues of $Q$ are different. Moreover, the matrix $Q$ is symmetric so its eigenvectors $e_i$ form an orthonormal basis. Let the corresponding real eigenvalues be $\lambda_i$. We can express $x_0$ in this base as $x_0 = \sum_i \alpha_i e_i$ for some $\alpha_i$. Using these notations we may rewrite the previous equation as

$$\frac{\sum_i (1 - \lambda_i)\alpha_i^2}{\sum_i \alpha_i^2} < \frac{24}{n^2}.$$

It is clear that $1 - \lambda_i^k < k(1 - \lambda_i)$ for any $\lambda_i \in [-1, 1]$, so it follows that

$$\frac{\sum_i (1 - \lambda_i^k)\alpha_i^2}{\sum_i \alpha_i^2} < \frac{24k}{n^2},$$

or with the original matrix notation

$$\frac{x_0(I - Q^k)x_0^T}{x_0 x_0^T} < \frac{24k}{n^2},$$

$$\frac{x_0 Q^k x_0^T}{x_0 x_0^T} > 1 - \frac{24k}{n^2}.$$

This is what we need for the part pointing in the $x_0$ direction, so let us now focus on the remainder.

Let us look at the orthogonal decomposition $x_0 Q^k = \alpha x_0 + y$, where $\alpha > 1 - 24k/n^2$ according to the previous estimate. The matrix $Q$ is non-expanding w.r.t. the $\|\cdot\|_2$ norm, so we have

$$\alpha^2 \|x_0\|_2^2 + \|y\|_2^2 \le \|x_0\|_2^2.$$

We need to transform this inequality to bound $\|y\|_1$. We can do this using the inequality of arithmetic and quadratic means:

$$\frac{\|y\|_1^2}{n} \le \|y\|_2^2 \le \|x_0\|_2^2 (1 - \alpha^2).$$

Here $\|x_0\|_2^2 = 1/(3n) + 8/(3n^2) < 2/n$ for $n \ge 2$. The final estimate is

$$\left\| \left( x_0 + \frac{\mathbf{1}}{n} \right) Q^k - \frac{\mathbf{1}}{n} \right\|_1 \ge \alpha \|x_0\|_1 - \|y\|_1 \ge \alpha \frac{1}{2} - \sqrt{2(1 - \alpha^2)}.$$

It is easy to verify that this is more than $5/12$ if $\alpha > 599/600$. We can ensure this whenever $k < n^2/15\,000$, so in the end we get that $c_1 = 1/15\,000$ is a sufficient choice for the lemma to be true. □

### *4.2. Non-reversible, but lazy chains*

As we outlined before, we want to relate our generic chain to a reversible one. We use the vector $x$ previously defined. Let us look at the following decomposition:

$$x(Q + rR)^k = xQ^k + \sum_{l=1}^{k} xQ^{l-1} rR(Q + rR)^{k-l}. \tag{5}$$

We know how the first term behaves, so we need to see that the other term is small. $Q + rR$ is non-expanding w.r.t. $\|\cdot\|_1$, so estimating $xQ^{l-1}rR$ is enough. If $r < c/n$ then there is some hope, as for $l = 1$ this vector has elements of size $8c/n^3$, so $\|xrR\|_1 \le 8c/n^2$, which is acceptable if we want to add up an order of $n^2$ of these.

We want a similar inequality for the other $l$, but for this we need the chain to be very lazy, which means $q_i \le 1/4$ for all $i$. We can ensure this by replacing $Q$ with $(3I + Q)/4$, but later we will have to deal with the problem to get back to the original $Q$.

To get a different view on this error term we may use the estimate

$$\|yR\|_1 = \sum_{i=0}^{n-1} |y_{i+1} - y_{i-1}| \le 2 \sum_{i=0}^{n-1} |y_{i+1} - y_i| =: 2V(y). \tag{6}$$

In other words we are measuring how much the coordinates of a vector vary as we go around the cycle. The following lemma is what we need to bound this.

**Lemma 10.** *Suppose $Q$ is as before and $q_i \le 1/4$ for all $i$. Using the previously defined $x$ and any $k \ge 0$,*

$$V(xQ^k) \le \frac{4}{n}.$$

**Proof.** The proof is cleaner if we assume that the coordinates of $xQ^k$ are different for each $k$. This is allowed by using Lemma 2.

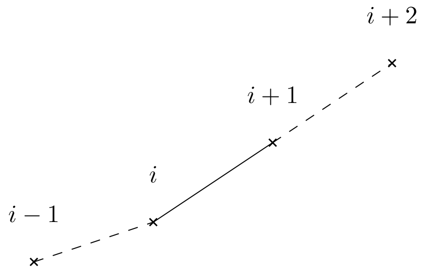


Fig. 2. Two non-peak nodes.

If we go around the cycle we see that the coordinates of $x$ consist of two monotone series, so there are only two local extrema. We call these *peaks*. The key thing is to show that this property remains as we multiply by $Q$. During the proof we will look at a few consecutive nodes and a single time step at once and find out how their ordering can change. We will do this until we cover all possibilities which can occur.

We mostly work by modifying weighted sums of some $y_i$ by exchanging one $y_i$ with a larger $y_j$. This way we maintain a sequence of inequalities to find out the new ordering.

One possibility is if there are two non-peak nodes after each other (see Fig. 2). This means the 4 nodes form a monotone sequence, e.g. $y_{i-1} < y_i < y_{i+1} < y_{i+2}$. In this case, we have the following:

$$\begin{aligned}(yQ)_i &= y_{i-1}q_i + y_i(1-q_i-q_{i+1}) + y_{i+1}q_{i+1}\\ &\leq y_i(1-q_{i+1}) + y_{i+1}q_{i+1}.\end{aligned}$$

Here is the only other type of step we use. This time we change the weights instead of the values. We increase the weight of the larger $y_{i+1}$ and decrease the weight of the smaller $y_i$. We use the assumption $q_{i+1} < 1/4$.

$$\begin{aligned}\cdots &< y_iq_{i+1} + y_{i+1}(1-q_{i+1})\\ &\leq y_iq_{i+1} + y_{i+1}(1-q_{i+1}-q_{i+2}) + y_{i+2}q_{i+2}\\ &= (yQ)_{i+1}.\end{aligned}$$

Consequently the ordering of the values at nodes $i$ and $i+1$ will remain the same.

The only other setting that occurs initially if there is a peak node between two non-peak nodes (see Fig. 3). Without the loss of generality we may assume they are ordered as $y_{i-2} > y_{i-1} > y_i < y_{i+1} < y_{i+2}$, and $y_{i-1} < y_{i+1}$. A similar claim works as in the previous case:

$$\begin{aligned}(yQ)_i &= y_{i-1}q_i + y_i(1-q_{i+1}-q_i) + y_{i+1}q_{i+1}\\ &\leq y_i(1-q_{i+1}-q_i) + y_{i+1}(q_{i+1}+q_i).\end{aligned}$$

Here we use $q_i < 1/4$ as in the previous case.

$$\begin{aligned}\cdots &< y_iq_{i+1} + y_{i+1}(1-q_{i+1})\\ &\leq y_iq_{i+1} + y_{i+1}(1-q_{i+1}-q_{i+2}) + y_{i+2}q_{i+2}\\ &= (yQ)_{i+1}.\end{aligned}$$

The ordering between node $i$ and $i+1$ remains the same, but it might change between node $i-1$ and $i$. In either case, the number of peak nodes will not increase, although their position might change.

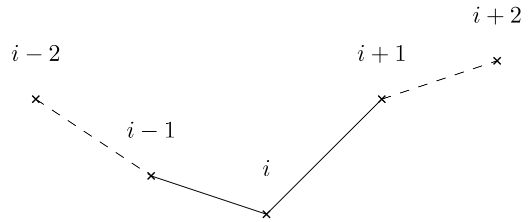


Fig. 3. Single peak node.

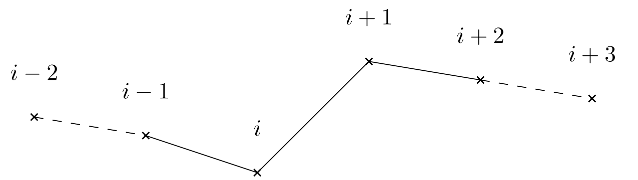


Fig. 4. Two peak nodes.

So far the only thing that could have happened is that these peaks moved around. After a few steps we might find a setting different from the previous two, namely when two peak nodes appear next to each other (see Fig. 4). We may assume they are ordered as $y_{i-1} > y_i < y_{i+1} > y_{i+2}$. As there are only 2 peak nodes, the sequence $y_i, y_{i-1}, \ldots, y_{i+1}$ is increasing, therefore $y_{i-1} < y_{i+1}$ and $y_i < y_{i+2}$. Now we have

$$\begin{aligned}(yQ)_i &= y_{i-1}q_i + y_i(1-q_{i+1}-q_i) + y_{i+1}q_{i+1}\\ &\le y_i(1-q_{i+1}-q_i) + y_{i+1}(q_{i+1}+q_i).\end{aligned}$$

We need $q_i < 1/4$ again, this time the condition is sharp.

$$\begin{aligned}&\le y_i(q_{i+1}+q_{i+2}) + y_{i+1}(1-q_{i+1}-q_{i+2})\\ &\le y_i q_{i+1} + y_{i+1}(1-q_{i+1}-q_{i+2}) + y_{i+2}q_{i+2}\\ &= (yQ)_{i+1}.\end{aligned}$$

This shows that at least the ordering in the middle will remain as it was. If any of the other two changes, it has the same effect as in the previous case, namely a peak node will become non-peak, and maybe the non-peak node after will become a peak node. So the number of peak nodes does not increase, therefore no other setting can occur.

We covered all possibilities, and the bottom line is that there are only two peaks for all $xQ^k$. Clearly one is a maximum, the other is a minimum, and for such vectors

$$V(y) = 2\left(\max_i y_i - \min_i y_i\right).$$

This difference does not increase in our case due to the fact that $Q$ is non-expanding w.r.t. $\|\cdot\|_\infty$. In the end $V(xQ^k)$ is at most its initial value, $4/n$. $\square$

Now we are ready to solve the lazy case.

**Lemma 11.** *Given a Markov chain on an $n$ node cycle consider the transition matrix $P = Q + rR$. Suppose $q_i \le 1/4$ for all $i$ and $0 \le r \le c/n$ for some fixed $c > 0$. Then there is* *a $c_2 > 0$ depending only on c such that for any $k \le c_2 n^2$ we have the bound*

$$\left\| xP^k - \frac{\mathbf{1}}{n} \right\|_1 > \frac{4}{12}.$$

*Consequently,*

$$t_{\mathrm{mix}}(P, 1/8) > c_2 n^2.$$

**Proof.** Consider the error introduced by the $rR$ terms in Eq. (5), use Eq. (6) and the previous lemma:

$$\begin{aligned}\left\| \sum_{l=1}^{k} xQ^{l-1} rR(Q+rR)^{k-l} \right\|_1 &\le \sum_{l=1}^{k} \left\| xQ^{l-1} rR(Q+rR)^{k-l} \right\|_1 \\ &\le r \sum_{l=1}^{k} \left\| xQ^{l-1} R \right\|_1 \le 2r \sum_{l=1}^{k} V(xQ^{l-1}) \le \frac{8rk}{n} \le \frac{8ck}{n^2}.\end{aligned}$$

If $k \le n^2/(100c)$, this error is at most $1/12$. We want to use Lemma 9 so fix $c_2 = \min(1/(100c), c_1)$, and choose $k \le c_2 n^2$. For such a $k$ we have

$$\begin{aligned}\left\| x(Q+rR)^k - \frac{\mathbf{1}}{n} \right\|_1 &\ge \left\| xQ^k - \frac{\mathbf{1}}{n} \right\|_1 - \left\| \sum_{l=1}^{k} xQ^{l-1} rR(Q+rR)^{k-l} \right\|_1 \\ &\ge \frac{5}{12} - \frac{1}{12} = \frac{4}{12}. \quad \square\end{aligned}$$

### *4.3. Relaxing laziness*

We need to transfer our conclusion to non-lazy chains. We use a binomial expansion to go back to the original $Q$.

$$x\left(\frac{3I}{4} + \frac{Q+rR}{4}\right)^k = \sum_{l=0}^{k} x \binom{k}{l} \frac{3^{k-l}}{4^k} (Q+rR)^l.$$

This allows us to form an inequality for the $l_1$ distances:

$$\left\| x\left(\frac{3I}{4} + \frac{Q+rR}{4}\right)^k - \frac{\mathbf{1}}{n} \right\|_1 \le \sum_{l=0}^{k} \binom{k}{l} \frac{3^{k-l}}{4^k} \left\| x(Q+rR)^l - \frac{\mathbf{1}}{n} \right\|_1.$$

We will carry through the following idea. Start with $k = \lfloor c_2 n^2 \rfloor$ as in Lemma 11. The right side is a weighted average of some $l_1$ distances. If the mixing time was very small for $Q + rR$, then these distances would be small for most of the terms. Then the average will be less than $4/12$ which we previously proved for the left hand side. This contradiction will prove our claim, and complete the theorem.

Suppose $t_{\mathrm{mix}}(Q + rR, 1/8) < k/8$. The $l_1$ distance is nonincreasing in $l$, so the terms are at most 1 and $1/4$ before and after $t_{\mathrm{mix}}(Q + rR, 1/8)$, respectively.

$$\sum_{l=0}^{k} \binom{k}{l} \frac{3^{k-l}}{4^k} \left\| x(Q+rR)^l - \frac{\mathbf{1}}{n} \right\|_1 \le \sum_{l=0}^{[k/8]-1} \binom{k}{l} \frac{3^{k-l}}{4^k} 1 + \sum_{l=[k/8]}^{k} \binom{k}{l} \frac{3^{k-l}}{4^k} \frac{1}{4}$$

$$\leq \frac{1}{4} + \frac{3}{4} P\left(Binom\left(k, \frac{1}{4}\right) < \frac{k}{8}\right).$$

This probability can be easily bounded e.g. by Chebyshev's inequality:

$$P\left(Binom\left(k, \frac{1}{4}\right) < \frac{k}{8}\right) \leq P\left(\left|Binom\left(k, \frac{1}{4}\right) - \frac{k}{4}\right| > \frac{k}{8}\right) \leq \frac{\frac{3k}{16}}{\frac{k^2}{64}} = \frac{12}{k}.$$

We can find an $n_0$ such that $n > n_0$ implies $k > 108$. In this case the probability is less than $1/9$, and the right hand side is strictly less than

$$\frac{1}{4} + \frac{3}{4 \cdot 9} = \frac{4}{12}.$$

This is the contradiction we were looking for.

In the end, let us choose $c' = \min(c_2, 1/n_0^2)$ so that the statement is also true for small $n$.

## 5. Looking forward

The theorem we proved sheds light on fundamental limitations of what can be achieved by a Markov Chain Monte Carlo method on the given topology. This result leads us to several related problems. The most natural question is to ask for a lower bound of the mixing time for other connectivity graphs. It is easy to answer this problem in the extreme cases. In the case of a tree, all chains will be reversible (assuming uniform stationary distribution), and known theory applies. For a complete graph mixing in 1 step is possible even without violating reversibility. On the other hand, we do not know much about what happens in between.

A special subset of connectivity graphs are those where a few extra edges are added to a cycle passing through all nodes. A special case is the example of the double cycle mentioned in the Introduction (from [9], Example 6.6) which can be viewed as a cycle with $2n$ nodes and $n$ extra edges. We know that a reversible chain with this connectivity graph has a mixing time of $\Omega(n^2)$ while a non-reversible chain can decrease this to $O(n)$. As another option, we can think of adding random edges. When using $\Omega(n)$ extra edges, we step into the territory of Small World Networks, and we may view our graph as coming from a slightly modified Watts–Strogatz model (see [11]). It is known that mixing time can be as low as $\log(n)$ (see [6,1]). This is a spectacular drop compared to the mixing time of the original ring. It raises the question of what can we achieve with a smaller number of extra edges.

Another direction to look forward is the problem of graph design: here one may want to find the fastest mixing chain satisfying specific constraints such as an upper bound on the edges of the connectivity graph, or a locality constraint. Note that general methods, such as the Metropolis–Hastings algorithm [7] do not give the fastest mixing chain for specific problems. Namely, if we want to sample from the uniform distribution, then it necessarily produces a reversible chain, and it does not exploit the possibility of using a non-reversible one. This alone shows that the problem of design deserves a closer look.

## Acknowledgments

I would like to express my thanks to my temporary supervisor Professor John Tsitsiklis for motivating me to investigate the subject matter of this paper, and for his guidance during my research.

I would also like to thank László Gerencsér and Gábor Tusnády for their clarifying questions and helpful comments.

This work was supported by the Hungarian–American Fulbright Commission and hosted by the Laboratory for Information and Decision Systems at the Massachusetts Institute of Technology.

## References

[1] A.J. Alireza Tahbaz-Salehi, Small world phenomenon, rapidly mixing Markov chains, and average consensus algorithms, in: 46th IEEE Conference on Decision and Control, 2007, pp. 276–281. doi:10.1109/CDC.2007.4434174.

[2] S. Boyd, P. Diaconis, P. Parrilo, L. Xiao, Fastest mixing Markov chain on graphs with symmetries, SIAM J. Optim. 20 (2) (2009) 792–819. doi:10.1137/070689413. URL: http://dx.doi.org/10.1137/070689413.

[3] S. Boyd, P. Diaconis, L. Xiao, Fastest mixing Markov chain on a graph, SIAM Rev. 46 (4) (2004) 667–689 (electronic) doi:10.1137/S0036144503423264. URL: http://dx.doi.org/10.1137/S0036144503423264.

[4] F. Chen, L. Lovász, I. Pak, Lifting Markov chains to speed up mixing, in: Annual ACM Symposium on Theory of Computing, Atlanta, GA, 1999, ACM, New York, 1999, pp. 275–281 (electronic).

[5] P. Diaconis, S. Holmes, R.M. Neal, Analysis of a nonreversible Markov chain sampler, Ann. Appl. Probab. 10 (3) (2000) 726–752. doi:10.1214/aoap/1019487508. URL: http://dx.doi.org/10.1214/aoap/1019487508.

[6] R. Durrett, Random Graph Dynamics, Cambridge University Press, 2006.

[7] W.K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1) (1970) 97–109. doi:10.1093/biomet/57.1.97. arXiv:http://biomet.oxfordjournals.org/cgi/reprint/57/1/97.pdf, URL: http://biomet.oxfordjournals.org/cgi/content/abstract/57/1/97.

[8] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (6) (1953) 1087–1092. doi:10.1063/1.1699114. URL: http://link.aip.org/link/?JCP/21/1087/1.

[9] R. Montenegro, P. Tetali, Mathematical aspects of mixing times in Markov chains, Found. Trends Theor. Comput. Sci. 1 (3) (2006) x+121.

[10] A. Olshevsky, J.N. Tsitsiklis, Convergence speed in distributed consensus and averaging, SIAM J. Control Optim. 48 (1) (2009) 33–55. doi:10.1137/060678324. URL: http://dx.doi.org/10.1137/060678324.

[11] D.J. Watts, S.H. Strogatz, Collective dynamics of ‘small-world’ network, Nature 393 (1998) 440–442. doi:10.1038/30918. URL: http://dx.doi.org/10.1038/30918.