# EFFECTS OF LONG RANGE CONNECTIONS ON MARKOV CHAIN MIXING TIMES*

BALÁZS GERENCSÉR†

**1. Introduction.** The performance of Markov Chain Monte Carlo (MCMC) methods strongly depend on the underlying Markov chain. We always try to make sure that it reaches its stationary distribution fast. For a fixed Markov chain, this means we have to prove the chain is near equilibrium after a certain number of steps. Sometimes we have some freedom in choosing the Markov chain, in this case we have the additional problem of finding the fastest one.

Distributed consensus and averaging methods can also be interpreted as Markov chains reaching the uniform distribution provided that this is the stationary distribution. For more details see Olshevsky, Tsitsiklis [13] and Boyd et al. [4].

Our scope is limited to discrete time, finite state space chains, with a unique stationary distribution $\pi$. The speed of approaching the stationary distribution can be quantified as the *mixing time*, defined as follows:

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

Here $\mathcal{P}(\mathcal{X})$ is the set of probability distributions on the state space $\mathcal{X}$, $\sigma^{(k)}$ is the distribution after $k$ steps with initial distribution $\sigma$, $\pi$ is the stationary distribution of the chain, and $\|.\|_{\mathrm{TV}}$ is the total variation norm defined for signed measures as:

$$\|\mu\|_{\mathrm{TV}} = \max_{A\subset\mathcal{X}} |\mu(A)|.$$

Sometimes we simply write $t_{\mathrm{mix}}$, assuming $\varepsilon = 1/4$.

A frequently used property of certain chains is *reversibility* (see e.g. Kelly [9]). A Markov chain is called reversible if starting from the stationary distribution, the probability of the consecutive pair $(i, j)$ is the same as the probability of the consecutive pair $(j, i)$. Formally:

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

where $p_{ij}$ is the transition probability from $i$ to $j$. In the rest of the paper we restrict our attention to the special case in which the stationary distribution is uniform. In this case the reversibility condition simplifies to $p_{ij} = p_{ji}$ for every pair $(i, j)$.

Sometimes we tweak the chain by changing transition probabilities but not the allowed transitions. For this purpose, let us define the connectivity graph of a Markov chain as follows: we use the states as nodes and include edges $(i, j)$ if either $p_{ij} > 0$ or $p_{ji} > 0$. We usually omit self loops from this graph.

In the next section we provide a brief overview of some speedup techniques. After that we give some details on our current work concerning speedup by adding a small number of long range edges to the connectivity graph.

*Funded by the COGANGS project of the European Union's Seventh Framework Programme (FP7-SME-2010-1-262017-COGANGS).

†HAS Rényi Institute

**2. Speedup techniques.** As a first step, we leave the connectivity graph of the chain intact and try to change the transition probabilities to decrease mixing time. If we require the chain to be reversible, there is a method by Boyd et al. [3] to formulate the question as a semi-definite programming (SDP) problem. Thus we may find the fastest reversible chain as long as we have an SDP solver fast enough for the size of the problem.

Often there is no good reason to require reversibility, and the question arises what we can gain by dropping this condition. Without going into details, one can show using conductance estimates that the mixing time might decrease to its square root for the best setups. This gain is indeed achieved in some cases as demonstrated in Example 6.6. by Montenegro and Tetali [11]. Here the connectivity graph is a pair of cycles interconnected at each node (imagine a hamster wheel). The mixing time is of the order of $n^2$ and $n$ for the best reversible and non-reversible chains, respectively. But this is not always the case, as shown by the author [8]:

THEOREM 2.1. *Consider a Markov chain on a cycle with n nodes. Suppose the stationary distribution is uniform. Then, for some global constant* $C > 0$*,*

$$t_{\text{mix}}(1/8) \geq Cn^2.$$

Note that the mixing time for the best reversible chain is also of the order of $n^2$.

Let us now turn our attention to other concepts where we allow the change of the connectivity graph in a certain way. One well known method is the idea of *lifting* which was introduced by Diaconis, Holmes and Neal in [6] and extended by Chen, Lovász and Pak [5]. The idea is to split each state into two or more new states and determine transition probabilities in an appropriate way so that the marginal of the new chain will behave like the original one. The additional structure allows more flexibility and once again, it can decrease mixing time up to the square root of the original.

The other type of modification we consider is the addition of some new edges to the connectivity graph. Let us start with a cycle with $n$ nodes. If we add $\approx cn$ random edges we suddenly arrive to Small World Networks (SWN), see the model of Newman et al. [12]. There is an intensive research activity on SWNs, including the study of random walks on them. The following result is due to Durrett [7], Addario-Berry and Lei [1] as it is roughly quoted here:

THEOREM 2.2. *Consider an n node graph from the model of Newman et al. For the symmetric random walk we have*

$$c_1 \log^2 n < t_{\text{mix}} < c_2 \log^2 n$$

*asymptotically almost surely as* $n \to \infty$*.*

This is a huge speed gain compared to the mixing time of $n^2$ for the cycle alone. At this point our goal is to investigate the options and effects of adding a lower, $o(n)$ number of extra edges. Our current results are presented in the following section.

**3. Few long range edges.** We start with a cycle with $n$ nodes and we want to add a small number of long range edges. In the beginning we were trying to design the graph using our intuitions, but it was quite annoying to realize how much better random graphs were compared to our “clever” designs.

So we focus on random long range edges. The target edge density of the added edges is $n^{-\alpha}$ for some parameter $\alpha \in (1, 2)$. Thus we expect $cn^{2-\alpha}$ extra edges. Let us introduce three slightly different models for choosing them.

M1: We add a random matching on the almost equidistant $\lceil n^{2-\alpha} \rceil$ nodes $\{[in^{\alpha-1}],\ 0 \leq i < n^{2-\alpha}\}$.

M2: From all possible long range edges we draw $\lceil n^{2-\alpha} \rceil$ randomly uniformly.

M3: For all possible long range edge we randomly decide to include it or not. Each edge is included independently with probability $n^{-\alpha}$.

In each case we assume the transition probabilities are chosen to achieve the fastest mixing while keeping the stationary distribution uniform. The models differ only slightly, but the results are different and depend on the techniques we can apply for them.

Sometimes we only consider the simple case of *homogeneous* chains when there are three common transition probabilities for long range edges, for clockwise and counter-clockwise transitions.

The first result is a simple lower bound.

PROPOSITION 3.1. *For M1, let us assume the nodes with long range edges are placed in an exactly equidistant way. Then for any homogeneous chain,*

$$t_{\text{mix}} \geq Cn^{2\alpha-2}.$$

*Proof.* Observe that we can "wind up" the chain around a cycle of $n^{\alpha-1}$ nodes so that long range edges become loop edges. Thus the result follows from Theorem 2.1. □

Let us define the *conductance* of a Markov chain. This is a quantity somehow measuring the worst bottleneck of the chain.

DEFINITION 3.2. *The conductance of a Markov chain on a state space $\mathcal{X}$ is*

$$\Phi = \min_{\emptyset \neq S \subsetneq \mathcal{X}} \Phi(S) = \min_{\emptyset \neq S \subsetneq \mathcal{X}} \frac{Q(S, S^C)}{\pi(S)\pi(S^C)} = \min_{\emptyset \neq S \subsetneq \mathcal{X}} \frac{\sum_{i \in S, j \in S^c} \pi_i p_{ij}}{\pi(S)\pi(S^C)}.$$

This quantity is strongly related to mixing times. The simplest universal bound is the following, see Borgs [2] for a proof.

PROPOSITION 3.3. *For any discrete time Markov chain we have the bound*

$$c\frac{1}{\Phi} \leq t_{\text{mix}}.$$

Moreover there are upper bounds on mixing time as well, let us cite the version of Lovász and Simonovits [10] which does not require the reversibility of the chain:

THEOREM 3.4. *For an irreducible, aperiodic Markov chains which is also lazy ( $p_{ii} \geq 1/2$ for all $i$) we have the following bounds for the mixing time:*

$$c_1\frac{1}{\Phi} \leq t_{\text{mix}} \leq c_2\frac{1}{\Phi^2}.$$

There is a square factor between the lower and upper bounds so we need additional tools if we look for a sharp estimate. Following the proof of Durrett [7] for SWN mixing time we can show the following:

THEOREM 3.5. *For M3 the conductance of the chain satisfies the following inequality asymptotically almost surely as $n \to \infty$:*

$$c\frac{n^{1-\alpha}}{\log n} < \Phi.$$

Let us sketch the concepts of the proof. We have to show that $\Phi(S)$ is large for every possible $S$. We define

$$B = \left\{ \emptyset \neq S \subsetneq \mathcal{X} \mid P\left(\Phi(S) < \frac{cn^{1-\alpha}}{\log n}\right) > 0 \right\},$$
$$k = |B|,$$
$$p = \max_{S \in B} P\left(\Phi(S) < \frac{cn^{1-\alpha}}{\log n}\right).$$

Now $kp$ is clearly an upper bound for the probability of violating the bound for $\Phi$. We only have to make sure it converges to 0 as $n$ grows. It is easy to see that $k$ increases and $p$ decreases exponentially, we need fine estimates to keep $kp$ under control.

PROPOSITION 3.6. *For M1, M2 and M3 the conductance of the chain satisfies the following inequality asymptotically almost surely as* $n \to \infty$*:*

$$\Phi < cn^{1-\alpha}.$$

*Proof.* Suppose for a moment that an instance of the random graph contains an arc $A$ at least $cn^{\alpha-1}$ long without a long range edge. Then we can bound the conductance:

$$\Phi = \min_{\emptyset \neq S \subsetneq \mathcal{X}} \frac{Q(S, S^C)}{\pi(S)\pi(S^C)} \leq \frac{Q(A, A^C)}{\pi(A)\pi(A^C)} \leq c\frac{2n^{-1}}{n^{\alpha-2}} = cn^{1-\alpha}.$$

Consequently it is enough to bound the probability of a random graph containing such an empty arc.

For M1 the arc $\left[1, [n^{1-\alpha}] - 1\right]$ is always appropriate.

For M2 the long range edges have a total number of $2[n^{1-\alpha}]$ endpoints distributed along $n$ nodes. Thus there must be a gap of at least $\frac{1}{2}n^{1-\alpha}$.

For M3 the number of edges $e$ follow a Binomial distribution. Using Chebyshev's inequality we get

$$P(e > n^{1-\alpha}) < P\left(\left|e - n^{-\alpha}\frac{n(n-3)}{2}\right| > \frac{1}{2}n^{2-\alpha}\right) < c\frac{n^2 n^{-\alpha}(1 - n^{-\alpha})}{n^{4-2\alpha}} = cn^{\alpha-2}.$$

Thus there are at most $n^{2-\alpha}$ edges asymptotically almost surely, in this case the reasoning for M2 applies.

□

Using Theorem 3.4 on the previous conductance bounds we arrive at the following corollary:

COROLLARY 3.7. *For M3 the mixing time of the chain satisfies the following inequality asymptotically almost surely:*

$$c_1 n^{\alpha-1} < t_{\text{mix}} < c_2 n^{2\alpha-2} \log^2 n.$$

*Moreover, if the chain is reversible we have*

$$c_1 n^{2\alpha-2} < t_{\text{mix}} < c_2 n^{2\alpha-2} \log^2 n$$

*asymptotically almost surely.*

*Proof.* We only need to show the lower bound for reversible chains. The others follow by combining Proposition 3.6 with Proposition 3.3 and Theorem 3.5 with Theorem 3.4. From the proof of Proposition 3.6 we know the random graph has an arc at least $cn^{\alpha-1}$ long without a long range edge asymptotically almost surely. Without going into details, mixing only within this portion of the chain needs at least $cn^{2\alpha-2}$ steps which provides the missing bound. ☐

For reversible chains the gap is reasonably tight. For non-reversible chains it is still unclear where the truth lies in between these bounds.

Still, we hope there is a considerable gain for non-reversible chains as shown in Figure 3.1. This is a plot of mixing times of homogeneous reversible and non-reversible chains on several graphs coming from M2, for $\alpha = 1.5$.

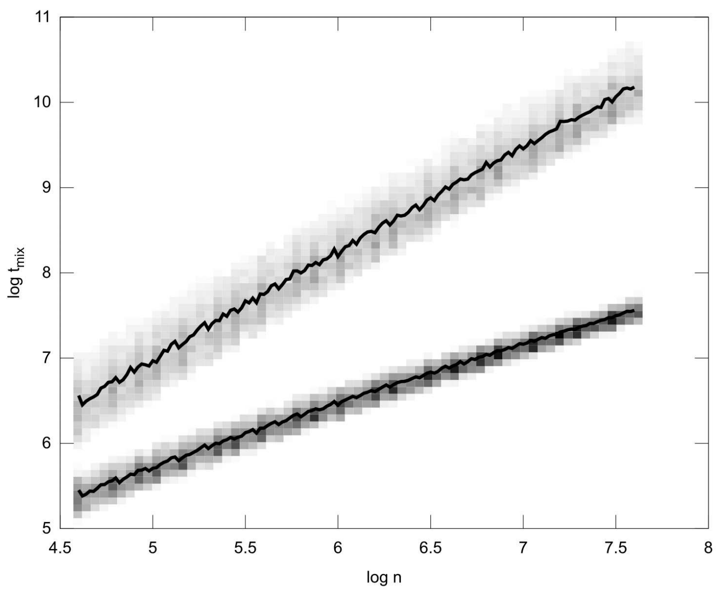


FIG. 3.1. *Log-log plot for mixing times of homogeneous M2 chains*

The sizes of the graphs change between 100 and 2000. We generated 222 random graphs for each size and calculated the mixing times for a reversible and a non-reversible chain on them. As we are not interested in the extremes, we discarded the lowest and highest 5% of mixing times for each size. The results for the remaining graphs are plotted as a histogram on a log-log scale. The upper cluster contains the data for the reversible chains, the lower for the non-reversible ones. The two lines are simply the averages.

It is clearly visible that non-reversible chains offer a significant speedup over reversible ones in this setting. We hope to quantify this gain in the future but at this point, we do not aim for a bold guess as $\log n$ and $n^\delta$ factors can be easily mistaken for each other on this scale.

On the other hand we may guess the mixing time for reversible chains is $n \log^\delta n$ based on Corollary 3.7. By looking for the best fit of the data we arrive at the estimate $\delta = 1.45$.

**4. Conclusions.** In this paper we presented the current state of an ongoing research project. We investigate two possibilities for decreasing the mixing time of Markov chains, one is by relaxing the reversibility condition, the other is by adding a small number of edges to the underlying connectivity graph. We have reasonably sharp bounds for reversible chains with added edges. For the non-reversible counterpart we have numerical evidence for the speedup, but determining the exact bounds of the mixing time is still open.

## REFERENCES


[1] L. Addario-Berry and T. Lei, *The mixing time of the Newman–small world*, ArXiv e-prints, (2012).

[2] C. Borgs, *Statistical physics expansion methods in combinatorics and computer science*, CBMS lecture notes (in preparation), (2004).

[3] Stephen Boyd, Persi Diaconis, and Lin Xiao, *Fastest mixing Markov chain on a graph*, SIAM Rev., 46 (2004), pp. 667–689 (electronic).

[4] Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah, *Randomized gossip algorithms*, IEEE Trans. Inform. Theory, 52 (2006), pp. 2508–2530.

[5] Fang Chen, László Lovász, and Igor 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).

[6] Persi Diaconis, Susan Holmes, and Radford M. Neal, *Analysis of a nonreversible Markov chain sampler*, Ann. Appl. Probab., 10 (2000), pp. 726–752.

[7] Rick Durrett, *Random Graph Dynamics*, Cambridge University Press, 2006.

[8] Balázs Gerencsér, *Markov chain mixing time on cycles*, Stochastic Processes and their Applications, 121 (2011), pp. 2553 – 2570.

[9] F.P. Kelly, *Reversibility and stochastic networks*, vol. 40, Wiley New York, 1979.

[10] László Lovász and Miklós Simonovits, *The mixing rate of Markov chains, an isoperimetric inequality, and computing the volume*, in 31st Annual Symposium on Foundations of Computer Science, Vol. I, II (St. Louis, MO, 1990), IEEE Comput. Soc. Press, Los Alamitos, CA, 1990, pp. 346–354.

[11] Ravi Montenegro and Prasad Tetali, *Mathematical aspects of mixing times in Markov chains*, Found. Trends Theor. Comput. Sci., 1 (2006), pp. x+121.

[12] M.E.J. Newman, C. Moore, and D.J. Watts, *Mean-field solution of the small-world network model*, Physical Review Letters, 84 (2000), pp. 3201–3204.

[13] Alex Olshevsky and John N. Tsitsiklis, *Convergence speed in distributed consensus and averaging*, SIAM Review, 53 (2011), pp. 747–772.