OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

Mixing Behavior of Uniformly Weighted and Gamma-Weighted Random Walks on Erdős–Rényi Graphs

Aksh Agarwal
aksh.agarwal0512@gmail.com

ABSTRACT

The study of Markov chains occupies a central position in probability theory, and the design of randomized algorithms. A Markov chain is a stochastic process defined by the property that the future state of the system depends only on its present state, known as memorylessness. As a chain continues to move, the fraction of time spent in each state of the chain begins to stabilize, to what is known as the "Steady State Distribution".

This paper compares the convergence of random walks with uniform-weighted and Gamma-normalized outgoing weights on matched directed Erdős–Rényi graphs. The paper analyzes 11 graphs from 50 to 100 states with uniform teleportation probability of $\tau = 0.001$, ensuring unique stationary distributions. For each chain, 100 Monte Carlo trials with 25 million walkers were used to estimate convergence time from uniform initial distribution with a total-variation tolerance of $\varepsilon = 0.01$, along with the worst-case mixing from every starting state. Uniform-start convergences took anywhere from 10-136 steps, while worst-case mixing was found to occur between 4,297-4592 steps. The uniform-start simulation estimates were within a single step of their exact references. Little consistent ordering was found between weighting schemes for uniform-start convergence across tested topologies.

KEYWORDS: *Markov chains; mixing time; directed Erdős–Rényi graphs; weighted random walks; Gamma-distributed weights; Monte Carlo simulation; total variation distance; teleportation.*

INTRODUCTION

Markov chains are an essential tool for modeling stochastic processes across disciplines. As a result, understanding mixing times is essential to their use in real-world modeling. In statistical physics, for example, Markov Chain Monte Carlo (MCMC) methods are the primary computational tool for sampling from Boltzmann distributions and studying phase transitions in models such as the Ising model. MCMC methods use stochastic transitions to sample equilibrium distributions and estimate physical quantities that are difficult to calculate directly. The Metropolis algorithm provides a foundational example of this approach for sampling configurations based on their Boltzmann probabilities (Metropolis et al., 1953). A

key question that arises with these models is how quickly the system's long-term behavior emerges. A Markov chain's mixing time addresses this by measuring how long it takes for the chain to become close to its long-run distribution, independent of its initial state.

Understanding which graph statistics may influence mixing time is therefore incredibly crucial to studying stochastic processes. Two basic features of a Markov chain are the structure of its state space and the probabilities that dictate movement between states. Both could affect the rate at which the chain approaches equilibrium.

This study investigates how uniform and Gamma-normalized outgoing transition weights affect convergence from uniform initial distributions on matched sparse directed Erdős–Rényi graphs, as well as how these times may compare to the worst-case mixing. The research question is whether different weighting schemes produce notable differences in uniform-start convergence on matched directed Erdős–Rényi graphs, how convergence varies with n, as well as how it compares with worst-case mixing. The paired comparison tests the hypothesis is that worst-case mixing is substantially slower than uniform-start convergence in the chosen rare-teleportation regime, and the secondary null-hypothesis that the Gamma and uniform weighted convergence-time differences are symmetric about zero

LITERATURE REVIEW

Mixing time & Convergence to Equilibrium

The mixing time of a finite ergodic Markov chain measures how long it takes for the distribution of the chain to approach steady state distribution. The steady state distribution π is determined when $\pi = \pi P$, where P is the transition matrix, and all the entries in π must sum to 1. If P is an $n \times n$ matrix, and $\pi = (\pi_1, \pi_2 \ldots \pi_n)$, then the value of π for a given entry must equal the sum probabilities of transitioning from all states from 1, up till j. Therefore: $\pi_j = \sum_{i=1}^{n} \pi_i P_{ij}$ for all $j \leq n$. The mixing time can be given where $t_{mix}(\varepsilon)$ is the smallest t such that $||\mu_t - \pi||_{TV} \leq \varepsilon$, where μ_t is the distribution of probabilities after t, TV is the total variation, and ε is the tolerance allowed for being 'close enough' to the steady state. A low mixing time indicates that the system converges fast, whereas a large one implies slow convergence.

The worst-case mixing time at tolerance ε is defined by the time such that the total-variation distance is lower than the allowed tolerance. Or:

$$t_{\text{mix}}(\varepsilon) = \min\left\{t : \max_x d_x(t) \leq \varepsilon\right\}$$

This definition demonstrates that mixing of a graph is dependent on both the transition probabilities, and the least favorable initial state. It also demonstrates why the number of states alone cannot determine the

rate of convergence. Levin and Peres (2017) present mixing-time theory as the study of how graph features can determine convergence to stationarity. Montenegro and Tetali (2006) similarly review several approaches, including spectral gaps, conductance and other isoperimetric quantities, logarithmic Sobolev and Nash inequalities (Diaconis and Saloff-Coste, 1996), entropy methods, and lower-bound techniques. These approaches establish that no single graph statistic is sufficient in determining the mixing time, however they indicate that varying certain statistics, with all other statistics remaining constant, can yield consistent changes.

A complication with theoretical mixing time approximations is the cutoff phenomenon. Diaconis (1996) described families of chains where the distance from stationarity remains large for an extended period and then falls to near zero over a relatively smaller time window. As a result, small changes in the total variation threshold may produce large apparent changes in the estimated mixing time over a wide observation interval. However, the cutoff phenomenon shouldn't necessarily be characterized as a failure of theoretical mixing-time methods, but rather a pattern that broad theoretical methods may not accurately account for.

Graph size, connectivity, & Erdős–Rényi structure

Previous work on random walks on Erdős–Rényi graphs shows that the effect of graph size is mediated by connectivity and local structure. Fountoulakis and Reed (2008) studied simple random walks on the largest component of a supercritical random graph, and concluded that different structural mechanisms control convergence in different degree regimes. In graphs with relatively small average degrees, local obstructions such as thin branches and poorly connected regions slowed down the random walk, increasing mixing times. As the average degree increased, these obstructions became less consequential, and global values such as graph diameter grew more influential. Thus, increasing the number of vertices wasn't shown to have a conclusive effect: rather the effect is dependent on how the edge probability and resulting degree distribution scale with n.

Benjamini, Kozma, and Wormald (2014) obtained a closely related result for the giant component of a supercritical Erdős–Rényi graph, showing that its total-variation mixing time is of order $(\log(n))^2$ in the sparse region they considered.

Many classical results relate to undirected graphs or random walks restricted to a giant connected component. Directed random graphs introduce additional complications because reachability need not be reciprocal and their stationary distributions need not be close to uniform. Bordenave, Caputo, and Salez (2018) studied random walks on sparse random directed graphs with prescribed in- and out-degree sequences. They established cutoff, described its window and limiting profile, and investigated the stationary distribution. Their results show that rapid and sharply concentrated convergence can also occur in nonreversible directed chains. However, it is worth noting that they rely on degree assumptions that differ slightly from finite directed G(n,p), graphs with dangling states.

Transition-weight heterogeneity

Boyd, Diaconis, and Xiao (2004) examined the fastest-mixing Markov chain problem. For a fixed, connected graph, they studied optimized transition probabilities to minimize the second-largest eigenvalue of the graph, therefore maximizing the spectral gap. The resulting formulation determined that two chains on the same graph could have substantially different mixing times owing to differing transition probabilities.

Sampling transition weights of a Markov chain from a Gamma-distribution has a precise probabilistic interpretation. Suppose a vertex has d outgoing edges and independent weights

$$W_1, \dots, W_d \sim \mathrm{Gamma}(\alpha, 1)$$

After row normalization (Devroye, 1986, p. 594).:

$$\left(\frac{W_1}{\sum_{j=1}^{d} W_j}, \dots, \frac{W_d}{\sum_{j=1}^{d} W_j} \right) \sim \mathrm{Dirichlet}(\alpha, \dots, \alpha)$$

Thus, when α=1, each row is a Dirichlet(1,…,1) probability vector conditional on its out-degree. The uniform-weight condition assigns exactly 1/d to every outgoing edge, whereas the Gamma-weight condition produces exchangeable but unequal probabilities. For a probability distribution on *d* outgoing neighbors, Shannon entropy is at most $\log_d$ with equality when every neighbor receives probability $1/d$ (Cover & Thomas, 2006). Thus, before teleportation is added, a Gamma-normalized row with d≥2 has strictly lower entropy almost surely. For d=1. both weighting schemes have entropy zero.

Chafaï (2010) investigated the Dirichlet Markov ensemble, wherein the rows of a dense random stochastic matrix are independent Dirichlet vectors. The paper established asymptotic behavior for the singular values and identified characteristic spectral behavior of such random Markov matrices. Bordenave, Caputo, and Chafaï (2012) subsequently proved a circular-law result for row-normalized random matrices. In the exponential case, which corresponds to Gamma shape parameter one, the nontrivial complex eigenvalues of the appropriately rescaled dense matrix become distributed over a disk in the complex plane.

Their results should nevertheless be applied cautiously. The Dirichlet Markov ensemble is dense, and each row assigns probability across essentially all states. The given model first generates a sparse Erdős–Rényi support and then places a Dirichlet vector only on the outgoing neighbors. Row dimension can therefore vary with out-degree, and a substantial number of rows may contain only one or two nonzero entries. Dense random-matrix asymptotics can neither directly, nor accurately predict the finite-size mixing times of these sparse chains.

Bordenave, Caputo, and Salez (2019) provide a closer theoretical study to sparse random transition weights. They studied nonreversible sparse Markov chains in random environments with probability mass concentrated on relatively few entries, and exchangeable entries within each row. They established cutoff at an “entropic time” proportional by the below relationship

$$t_{\text{mix}} \propto \frac{\log n}{H}$$

where H is the average row entropy. This result suggests a means through which Gamma weighting could influence mixing. Unequal weights reduce the effective number of plausible next states and therefore lower the entropy generated by each step. Under this model, lower average row entropy leads to a later entropic mixing scale, and a greater mixing time. This should, however, be considered as a theoretical assumption rather than a direct prediction.

The literature therefore doesn't provide a universal conclusion that greater transition-weight heterogeneity must either increase or decrease mixing time. Unequal weights could slow mixing by concentrating probability on a small number of edges, or by creating stationary "bottlenecks". A relevant way to alter convergence is through lifting, wherein the state space is enlarged. Through this, suitable constructions can speed up mixing (Chen et al., 1999). The resulting outcome depends highly on row entropy, reciprocity, conductance, and the alignment of directed paths.

Teleportation and empirical estimation of mixing time

Teleportation ensures irreducibility and aperiodicity by replacing the original matrix with a mixture such as

$$P_\tau = (1-\tau)P + \tau J$$

where every row of J is the same strictly positive resampling distribution (Langville and Meyer, 2004). However, teleportation is not mathematically neutral. Caputo and Quattropani (2021) studied Google PageRank-type walks on sparse random digraphs and identified three convergence regimes. When the refresh probability is much smaller than the inverse native mixing scale, convergence is largely determined by the underlying walk. When refreshing is much more frequent, convergence is dominated by the refresh mechanism and has an approximately exponential form. An intermediate regime combines the two behaviors. This models the behavior of standard Markov chains, thus, teleportation can alter the shape and scale of convergence rather than merely guaranteeing the existence of a stationary distribution.

If a state has been assigned a self-loop due to a lack of outgoing edges, movement away from that state occurs primarily through teleportation. These states can therefore dominate worst-case mixing-time statistics.

Wolfer and Kontorovich (2019) address the problem of estimating mixing time from one trajectory, whereas the present method can simulate many independent walks from chosen initial states. Their work demonstrates why empirical claims about nonreversible mixing are reliant on quantification of uncertainty, and why pseudo-spectral quantities are preferable to direct eigenvalue estimates. In the present design, Monte Carlo variation affects the estimated total-variation curve and therefore the first threshold-crossing time. This is crucial near a cutoff or when the curve approaches ε=0.01 slowly. Confidence intervals or repeated batches of simulated walks would thus make it possible to distinguish graph-to-graph variation from simulation error.

METHODOLOGY

Directed Erdős–Rényi graphs were generated under the G(n, p) model, with the edge probability fixed at p = 0.03, and n = 50, 55, …, 100. One topology was generated at each size using graph seeds 42, 43… 52.

Two Markov chains were constructed from each topology. In the uniform-weight condition, probability was divided equally among all outgoing edges of a node. For a node i with out-degree d_i^+, the transition probabilities were defined as follows:

$$P_{ij} = \frac{1}{d_i^+} \, (j \in N^+(i)), \qquad P_{ij} = 0 \, (j \notin N^+(i))$$

In the Gamma-weighted condition, a positive value was independently sampled for each outgoing edge from a Gamma distribution with shape parameter α = 1 and scale parameter 1. The sampled values were normalized within each row:

$$P_{ij} = \frac{Y_{ij}}{\sum_{\ell \in N^+(i)} Y_{i\ell}}, \qquad Y_{ij} \sim \text{Gamma}(1, 1)$$

Nodes with no outgoing edges were assigned a self-loop of probability one. Each resulting graph was saved in GraphML format and reused in the simulation and theoretical analyses.

Uniform teleportation was introduced to ensure irreducibility and aperiodicity. The adjusted transition matrix was

$$P_\tau = (1 - \tau)P + \tau J$$

where τ = 0.001 and every entry of J was equal to 1/n. The value of τ leaves a mean interval of 1000 steps between teleportation events, allowing for long periods of movement governed by the native graph, while ensuring irreducibility and aperiodicity.

The stationary distribution π was computed by solving $\left[I - (1 - \tau)P^{\mathsf{T}}\right] \pi^{\mathsf{T}} = \tau u^{\mathsf{T}}$, where u = (1/n, …, 1/n). The solution was normalized and then checked for nonnegative entries as well as the stationarity condition. Exact reference distributions were obtained through repeated multiplication by P_τ. Beginning with the identity matrix yields the distribution from each individual state as its rows. A separate calculation begins from u. These yield the exact uniform-start convergence time and worst-case mixing time, with states evaluated at every graph size.

Simulation-Based Mixing Time

For each selected weighted chain, 100 independent trials were run with 25,000,000 independent random walkers per trial. Each walker's initial state was sampled independently from the uniform distribution. Walker positions were represented through occupancy counts . At each step, multinomial draws were used

to move remaining walkers along their native edges, while binomial draws selected teleporting walkers. A further multinomial draw was used to assign uniform destinations to teleporting walkers. The empirical distribution was obtained by dividing the state occupancies by N. The positions of the walks were used to estimate the state distribution. Its distance from stationarity was measured by total variation distance:

$$d_{\mathrm{TV}}(\hat{\mu}_t, \pi) = \frac{1}{2} \sum_{j=1}^{n} |\hat{\mu}_t(j) - \pi(j)|$$

A common convergence tolerance of ε = 0.01 was used across both weighting schemes. This provides a stricter accuracy criterion compared to traditional thresholds used in several standard definition of mixing times (Levin and Peres, 2017, Chapter 4). Since total variation is the largest absolute difference in probability assigned to any set of states, the chosen tolerance requires each such probability to be within one percentage point of its stationary value.

Sampling error was bounded simultaneously over the complete experiment. With K = 22 × 105 × 6,001 and δ = 0.05, the total variation error radius was $r_n = \sqrt{\frac{n \log 2 + \log(K/\delta)}{2N}} + \eta$. Here, $\eta = \frac{\left\| \hat{\pi} P_\tau - \hat{\pi} \right\|_1}{2\tau}$ bounds error in the stationary reference. The concentration bound and a union bound provide at least 95% simultaneous coverage for all chain, trial and time comparisons (Weissman et al., 2003). The sampling part of the radius ranged from 0.001040 to 0.001332.

For each run, the lower mixing-time endpoint was one plus the last observed time at which $\hat{d}(t) - r_n > \varepsilon$. t If no such time was recorded the lower endpoint was marked as 0. The upper endpoint was the first time at which $\hat{d}(t) + r_n \leq \varepsilon$. These endpoints bracket the true crossing time on the simultaneous coverage event. All runs in the experiment obtained upper endpoints, thus nothing was recorded for missing upper endpoints.

In order to check the worst-case timescale, 5 additional independent trials per chain began at a fixed state attaining its exact worst-case mixing time. Ties in this scenario were resolved by the smallest state index. These trials were compared with that state's exact curve and the same uncertainty procedure was applied uniformly. Stationary sampling noise was also checked using 1000 independent multinomial batches per chain.

Statistical comparison used 11 matched topologies as experimental units. Exact reference times were analyzed through two-sided Wilcoxon signed-rank tests for weighting comparisons between paired ranks based on weight, and Spearman rank correlations were used to derive associations with n. Signed-rank p-values were computed using exhaustive sign enumeration, with average ranks for tied absolute differences and zero differences omitted. Trend p-values were computed from all 11! assignments of the observed ranks to graph sizes. Holm adjustment was also applied jointly to the size associations for

uniform-start convergence time under both weighting schemes for worst-case and uniform distribution starting states. These analyses were exploratory in nature, since the signed-rank null assumes independent paired differences are symmetric about zero, and the trend permutation null assumes that there is exchangeability across graph sizes under no association.

Simulations and relevant statistical analyses were implemented in Python 3.12.14, using NumPy 2.3 for numerical calculations as well as random sampling. SciPy 1.17.0 was used for sparse-matrix operations, and statistical routines. Graph topologies and gamma weights were generated using Python's standard Random module. The code

RESULTS

Mixing times were collected across 100 simulation trials for each chain type and graph size. The results consist of Gamma-weighted chains for n=50 through n=100, and uniform-weighted chains. For each condition, the mean, median, standard deviation, minimum, maximum, 10th percentile, 90th percentile, and coefficient of variation were determined. Tables 1 and 2 summarize the uniform-start first crossings, comparing them with the determined exact references. Table 3 provides additional insight into the worst-case references and simulations.

n	Exact mixing time	MC mean	MC median	MC SD	MC p10	MC p90	MC min–max	95% bracket
50	27	27.61	28	0.49	27	28	27–28	[27, 29]
55	83	83.21	83	0.41	83	84	83–84	[81, 86]
60	14	14	14	0	14	14	14–14	[14, 14]
65	14	14	14	0	14	14	14–14	[14, 14]
70	53	53	53	0	53	53	53–53	[52, 54]
75	53	53.04	53	0.2	53	53	53–54	[52, 55]
80	33	33	33	0	33	33	33–33	[33, 34]
85	80	80	80	0	80	80	80–80	[78, 82]
90	121	121	121	0	121	121	121–121	[118, 125]
95	76	75.82	76	0.39	75	76	75–76	[74, 78]
100	136	135.74	136	0.44	135	136	135–136	[132, 139]

Table 1: Gamma-Weighted Chain Summary

n	Exact Mixing Time	Uniform MC mean	Uniform MC median	Uniform MC SD	Uniform MC p10	MC p90	MC min–max	95% bracket
50	10	10	10	0	10	10	10–10	[9, 10]
55	20	20	20	0	20	20	20–20	[20, 21]
60	19	19	19	0	19	19	19–19	[19, 20]
65	12	12	12	0	12	12	12–12	[12, 12]
70	48	48	48	0	48	48	48–48	[47, 50]
75	40	40	40	0	40	40	40–40	[39, 41]
80	34	34	34	0	34	34	34–34	[33, 35]
85	106	106.16	106	0.37	106	107	106–107	[104, 109]
90	94	94	94	0	94	94	94–94	[91, 96]
95	108	107.94	108	0.24	108	108	107–108	[105, 111]
100	134	134	134	0	134	134	134–134	[131, 138]

Table 2: Uniform-Weighted Chain Summary

n	Sinks	Uniform Exact Worst Time	Uniform MC median	Uniform 95% bracket	Gamma exact worst time	Gamma MC median	Gamma 95% bracket
50	15	4583	4591	[4493, 4697]	4583	4584	[4491, 4692]
55	12	4585	4588	[4486, 4696]	4585	4587	[4489, 4698]
60	8	4587	4590	[4486, 4701]	4587	4591	[4493, 4705]
65	11	4588	4591	[4488, 4710]	4588	4591	[4488, 4710]
70	9	4589	4592	[4485, 4715]	4589	4595	[4486, 4716]
75	9	4590	4594	[4482, 4715]	4590	4594	[4487, 4722]

80	11	4591	4598	[4482, 4726]	4591	4596	[4483, 4726]
85	5	4585	4591	[4479, 4719]	4583	4593	[4477, 4722]
90	5	4592	4603	[4486, 4742]	4592	4602	[4487, 4740]
95	7	4560	4568	[4453, 4699]	4579	4587	[4469, 4724]
100	3	4297	4317	[4191, 4430]	4356	4364	[4245, 4502]

Table 3: Worst-Case Times

DISCUSSION

The results indicate a strong distinction between different initial distributions. A population initialized uniformly over the graph, naturally appears to converge significantly sooner than one initialized at least favorable states for all trials.

Comparison between uniform-weighted and Gamma-weighted chains use each shared topology as a paired observation. For uniform-start convergence, the median paired difference (The difference between gamma and uniform weighted mixing time), was 2 steps, with a range of -32 to 63 steps. Gamma weighting yielded shorter convergence times in four of the 11 pairs, and longer times in 7. A two-sided Wilcoxon signed-rank test gave $W = 23.5$ and $p = 0.422$, thus the paired analysis didn't provide sufficient evidence of a systematic difference between either weighting scheme. This result neither establishes statistical equivalence, nor rules out effects of weighting.

For worst-case mixing, the median paired difference was zero, with weighting schemes giving identical exact worst mixing time values at 8 graph sizes. Gamma weighting was faster by 2 steps at $n = 85$, and slower by 19 and 59 steps at $n = 95$ and 100 respectively. The signed-rank test yielded $W = 1.0$ and $p = 0.500$. Only three nonzero pairs contributed to this test, however, so the evidence for a general weighting effect on worst-case mixing is limited in nature.

The association between graph size and uniform-start convergence was positive for uniform weight, with a Spearman rank correlation coefficient of $\rho = 0.918$ (exact permutation $p = 0.000190$; Holm-adjusted $p = 0.00114$). For Gamma-weighted chains, the association was numerically weaker with $\rho = 0.630$, (exact permutation $p = 0.0421$; Holm-adjusted $p = 0.210$) and did not meet the 0.05 threshold after adjustment. Neither worst-case series demonstrated evidence of a consistent size association ($\rho = -0.073$ for uniform weights and -0.132 for Gamma-weighted chains; both Holm-adjusted $p = 1.000$). These results

demonstrate an upward correlation for uniform-start convergence under uniform weighting within the sampled series, but offer no conclusive backing for a trend on both weighting schemes.

CONCLUSION

The results show a positive association between n and uniform-start convergence time for uniform-weighted chains, although the relationship is not monotonic. Large differences between neighboring graph sizes, along with strongly right-skewed distributions in several cases, indicate that local graph structure has a significant influence on convergence.

The comparison between weighting schemes also produced little consistent ordering. Gamma weighting led to faster mixing for some graph sizes and slower mixing for others. This indicates that transition-weight heterogeneity in transition probabilities does not have a universal effect on mixing time. Rather, its impact appears to depend on how assigned weights interact with other features of underlying topology. In a few cases, uneven probabilities may create bottlenecks or concentrate probability along a limited number of routes, while in others they may allow probability to move efficiently through the graph.

The results also demonstrate that graph size cannot be considered independently of connectivity in this model. Given that the probability was held constant during formation of initial Erdos-Renyi graphs, increasing n also increased expected out-degree.

In all, the study indicates that mixing time in sparse directed Markov chains is determined by interaction between graph size, connectivity, local structure, and transition-weight distribution, as opposed to uniformly or consistently by any single property alone.

A potential extension could be to hold expected degree constant while varying n, and separately vary the Gamma distribution or teleportation probability. This may allow the individual effects of graph size, transition-weight heterogeneity, and connectivity to be isolated more clearly. Additionally, the graph seeds could be varied, and the experiment conducted on a larger number of graph topologies.

REPRODUCIBILITY

The relevant code is accessible at https://github.com/akshagarwal12/GammaMarkovChains

REFERENCES

Benjamini, Itai, et al. “The Mixing Time of the Giant Component of a Random Graph.” Random Structures & Algorithms, vol. 45, no. 3, 2014, pp. 383–407, doi:10.1002/rsa.20539.

Bordenave, Charles, et al. "Circular Law Theorem for Random Markov Matrices." Probability Theory and Related Fields, vol. 152, nos. 3–4, 2012, pp. 751–79, doi:10.1007/s00440-010-0336-1.

Bordenave, Charles, et al. "Cutoff at the 'Entropic Time' for Sparse Markov Chains." Probability Theory and Related Fields, vol. 173, nos. 1–2, 2019, pp. 261–92, doi:10.1007/s00440-018-0834-0.

---. "Random Walk on Sparse Random Digraphs." Probability Theory and Related Fields, vol. 170, nos. 3–4, 2018, pp. 933–60, doi:10.1007/s00440-017-0796-7.

Boyd, Stephen, et al. "Fastest Mixing Markov Chain on a Graph." SIAM Review, vol. 46, no. 4, 2004, pp. 667–89, doi:10.1137/S0036144503423264.

Caputo, Pietro, and Matteo Quattropani. "Mixing Time of PageRank Surfers on Sparse Random Digraphs." Random Structures & Algorithms, vol. 59, no. 3, 2021, pp. 376–406, doi:10.1002/rsa.21009.
Chafaï, Djalil. "The Dirichlet Markov Ensemble." Journal of Multivariate Analysis, vol. 101, no. 3, 2010, pp. 555–67, doi:10.1016/j.jmva.2009.10.013.

Chen, Fang, et al. "Lifting Markov Chains to Speed Up Mixing." Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing, Association for Computing Machinery, 1999, pp. 275–81, doi:10.1145/301250.301315.

Cover, Thomas M., and Joy A. Thomas. Elements of Information Theory. 2nd ed., Wiley, 2006, doi:10.1002/047174882X.

Devroye, Luc. Non-Uniform Random Variate Generation. Springer-Verlag, 1986, doi:10.1007/978-1-4613-8643-8.

Diaconis, Persi. "The Cutoff Phenomenon in Finite Markov Chains." Proceedings of the National Academy of Sciences of the United States of America, vol. 93, no. 4, 1996, pp. 1659–64, doi:10.1073/pnas.93.4.1659.

Diaconis, Persi, and Laurent Saloff-Coste. "Logarithmic Sobolev Inequalities for Finite Markov Chains." The Annals of Applied Probability, vol. 6, no. 3, 1996, pp. 695–750, doi:10.1214/aoap/1034968224.

Fountoulakis, Nikolaos, and Bruce A. Reed. "The Evolution of the Mixing Rate of a Simple Random Walk on the Giant Component of a Random Graph." Random Structures & Algorithms, vol. 33, no. 1, 2008, pp. 68–86, doi:10.1002/rsa.20210.

Langville, Amy N., and Carl D. Meyer. "Deeper Inside PageRank." Internet Mathematics, vol. 1, no. 3, 2004, pp. 335–80, doi:10.1080/15427951.2004.10129091.

Levin, David A., and Yuval Peres. Markov Chains and Mixing Times. With contributions by Elizabeth L. Wilmer, 2nd ed., American Mathematical Society, 2017, doi:10.1090/mbk/107.

Metropolis, Nicholas, et al. "Equation of State Calculations by Fast Computing Machines." The Journal of Chemical Physics, vol. 21, no. 6, 1953, pp. 1087–92, doi:10.1063/1.1699114.

Montenegro, Ravi, and Prasad Tetali. "Mathematical Aspects of Mixing Times in Markov Chains." Foundations and Trends in Theoretical Computer Science, vol. 1, no. 3, 2006, pp. 237–354, doi:10.1561/0400000003.

Weissman, Tsachy, et al. Inequalities for the L_1 Deviation of the Empirical Distribution. Technical report HPL-2003-97R1, Hewlett-Packard Laboratories, 2003, shiftleft.com/mirrors/www.hpl.hp.com/techreports/2003/HPL-2003-97R1.pdf.

Wolfer, Geoffrey, and Aryeh Kontorovich. "Estimating the Mixing Time of Ergodic Markov Chains." Proceedings of the Thirty-Second Conference on Learning Theory, edited by Alina Beygelzimer and Daniel Hsu, vol. 99, PMLR, 2019, pp. 3120–59, proceedings.mlr.press/v99/wolfer19a.html.