23rd International Symposium on Mathematical Theory of Networks and Systems
Hong Kong University of Science and Technology, Hong Kong, July 16-20, 2018

Mixing rates for the cycle with structured additional wiring

Balázs Gerencsér[1], Julien M. Hendrickx[2]

***Abstract*— We analyze the mixing rate of a class of Markov chains where two features together help significantly speed up mixing: first, a cycle graph is enriched with extra edges in a randomized but structured way. Second, the Markov chain on the graph is taken to be asymmetric. We show a bound on the mixing rate of these Markov chains and demonstrate that both these features contribute to the outcome.**

***Index Terms*— mixing rate, random graph, drift, consensus**
MSC2010: 60J10, 05C80, 37A25

I. Introduction

Markov chains appear in various places in applications as a fundamental underlying building element. For instance, Markov chain Monte Carlo is a widespread tool for sampling from a wide variety of distributions or for approximating integrals, see Jerrum [1], Diaconis [2]. Another application is average consensus, where the goal is to compute the average of initial values appearing at a multitude of nodes connected along a network (which could correspond to measurements, opinions, local computation outputs, etc.), see Olshevky and Tsitsiklis [3], Olfati-Saber et al. [4]. The list is far from complete.

In all the cases, there is a natural correspondence between the efficiency of the particular application and the mixing properties of the underlying Markov chain. Therefore there is a high demand for better understanding the mixing behavior of various Markov chains not only for their mathematical interest, but also for their application significance.

Currently our focus is on the consensus relation. This translates to Markov chains that have uniform stationary distribution. An important feature is that, beside the analysis of a specific system, sometimes there is some freedom available in designing the network. This can involve the possibility of choosing the interaction strengths but sometimes also modifying the network structure itself. One can then use this freedom to amplify the mixing effect, and hence speed up the process of computing and agreeing on an average, i.e. reaching a consensus. Determining how to do this precisely is a challenging question.

In our current setup, we initialize with a connection network that is as sparse and homogeneous as possible: a cycle. This is modified first by adding a stronger interconnection structure according to a pattern to a small number of "hubs". Note that this approach is related conceptually to the idea of Csermely et al. [5] whose theory describes that a wide variety of real-world networks have two types of nodes, some in the "core" with strong connectivity structure and some in the "periphery" with lower number of edges and relying on the "core" in being well-connected to the network. This is what happens in our construction as well: the rich interconnection of the "hubs" is paired with the loose structure of the cycle containing the overwhelming portion of nodes. Besides this type of network structure, a difference considering the weights is present in our case: a drift is also introduced to the cycle, that is, all clockwise weights are increased and counter-clockwise weights decreased by the same amount. This additional modification may seem counter-intuitive at first, but the resulting asymmetry will play an important role for the mixing to be efficient.

We present a generalization of our previous work [6] where we first demonstrated that a fast rate towards consensus is achievable with this type of construction. However, in that paper all-to-all connections were assumed among the hubs, now we loosen these constraints.

II. Main result

We now describe formally the graph model and the Markov chain together with the related definitions to be able to state precisely the theorem about mixing efficiency.

Definition 1: Given $k \leq n \in \mathbb{Z}^+$ and a doubly stochastic $k \times k$ matrix A we define the randomized Markov chain $P_n(A)$ on n nodes as follows. Starting from a cycle graph on n nodes, we randomly uniformly and independently select k different edges which we remove. For the ith remaining arc, $1 \leq i \leq k$, we mark the clockwise endpoint as a_i and the other end as b_i. We introduce transitions from b_i to a_j according to the matrix A. Within any arc we set transition probabilities to 1 along the arc all the way from a_i to b_i. An example of such a Markov chain is demonstrated in Fig. 1.

We focus on the mixing behavior of the resulting (random) Markov chain which we measure by the following quantity:

Definition 2: For a Markov chain with transition matrix P we define the *mixing rate* λ as

$$\lambda = \min \{1 - |\mu| \ : \ \mu \neq 1 \text{ is an eigenvalue of } P\}.$$

Using this scheme, we want to form statements for fixed patterns, but when the size of the graph increases to infinity.

[1]B. Gerencsér is with MTA Alfréd Rényi Institute of Mathematics, Hungary and ELTE Eötvös Loránd University, Department of Probability Theory and Statistics, Hungary. His work is supported by NKFIH (National Research, Development and Innovation Office) grant PD 121107 and KH 126505. gerencser.balazs@renyi.mta.hu

[2]J. M. Hendrickx is with ICTEAM Institute, Université catholique de Louvain, Belgium. His work is supported by the DYSCO Network (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian Federal Science Policy Office, and by the Concerted Research Action (ARC) of the French Community of Belgium. julien.hendrickx@uclouvain.be

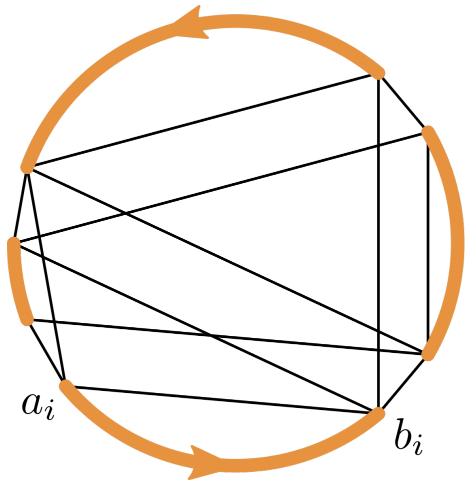


Fig. 1: Example for $P_n(A)$ (see Definition 1). Transitions along the orange arcs happen with probability one, transitions on the black lines are distributed according to the matrix A.

To formalize this, choose a *pattern* matrix B (which we will assume to be doubly stochastic and ergodic). Then we define $B^{*m} = \frac{1}{m} B \times K_m$ where K_m is the m-clique, that is, we take m copies of each node of B and connect two nodes the same way as their preimages were connected, then normalize properly. For a demonstration of this transformation see Fig. 2. With all the terminology introduced, the following theorem holds.

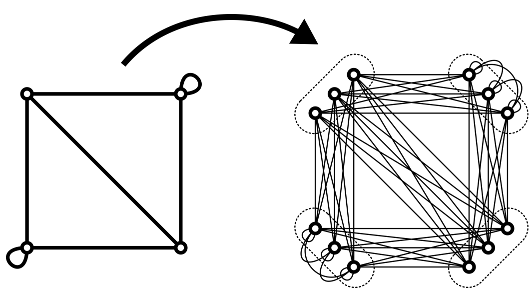


Fig. 2: Example for the scaling of patterns: for the transition matrix B of a 4 node Markov chain represented on the left we see the result of B^{*3} on the right (the exact numerical values are not shown).

Theorem 1: Given are a doubly stochastic ergodic matrix B and some constants $1 < \rho_l < \rho_u < \infty$. Assume $n, k \to \infty$ while $\rho_l < \log n / \log k < \rho_u$. Then for any $\gamma > 4$ asymptotically almost surely (a.a.s.) we have the following bound on the mixing rate of $P_n(B^{*k})$:

$$\lambda > \frac{k}{n \log^{\gamma} k}.$$

Simply speaking, the determining factor for the mixing rate is the average arc length $\frac{n}{k|B|}$ with some extra logarithmic factors (the constant size of B for the arc length is comparably negligible).

III. Comparison and numerical results

We first relate Theorem 1 to alternative situations, where the Markov chain is slightly different.

Most importantly, one could consider a version without drift, when a certain symmetric transition scheme is present along the arcs. Observe that locally within an arc we get a symmetric random walk. If a Markov chain is launched from the center of any arc, the escape time from that arc is quadratic in the length, therefore by standard tools [7] one can see that

$$\lambda < C \frac{k^2}{n^2}$$

would hold for some constant $C > 0$. This can be easily strengthened with a logarithmic factor a.a.s. This confirms that the drift indeed plays a key role in the high mixing rate achieved before.

The other natural modification is about choosing the interconnection structure among the hubs with more freedom, this is an open question in general to the best of our knowledge. Note that some extra edges are necessary, only the cycle itself gives rates that decrease quadratically in the length for any transition probabilities (even if non-homogeneous), see [8], [9] for a detailed discussion.

For specific examples we present numerical simulations for two different pattern matrices. In Fig. 3 we use the B already presented in Fig. 2, then in Fig. 4 we use a slightly larger and less interconnected alternative. In both cases we scale the pattern matrix so that $|B^{*k}| \approx \sqrt{n}$.

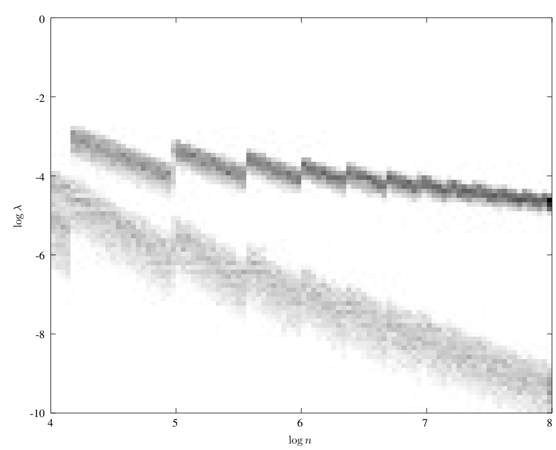
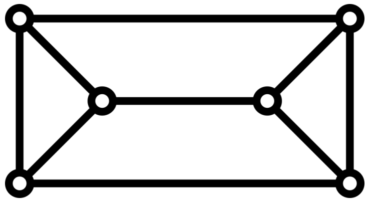


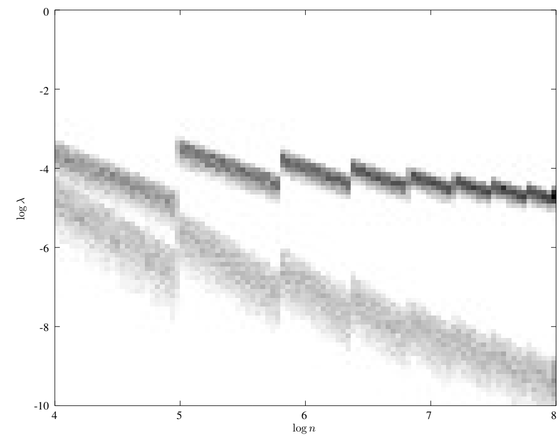
Fig. 3: Histograms for the mixing rates λ for the randomized Markov chains $P_n(B^{*k})$, using the pattern matrix in Fig. 2, scaled to have $\approx \sqrt{n}$ nodes.

Both figures are based on more than 30.000 randomly generated Markov chains with n ranging from 54 to 2980. As we are interested in typical behaviors of these randomized Markov chains, we discarded the top and bottom 5% of the results for every n considered.

In both cases we present a log-log histogram showing the decrease of λ as the node count n increases. The histogram presents the simulation results for the non-reversible and



(a) Graph representing the pattern matrix B (edges present for non-zero entries)



(b) Log-log histogram of the mixing rates for increasing node counts

Fig. 4: Simulations for mixing rates of random Markov chains chains $P_n(B^{*k})$ using the pattern matrix shown in (a), scaled to have $\approx \sqrt{n}$ nodes.

reversible Markov chain and we do observe the strong separation predicted by the theoretical results. The stripe on the top presents λ for the non-reversible Markov chains while the bottom one corresponds to the reversible ones. There is an interesting discontinuous behavior visible on both histograms. This can be traced back to the occasions when we step up with the scaling of the pattern matrices due to the node count increase. Nevertheless, for larger sizes the asymptotic behaviour starts to become the determining component, providing a $1/2$ factor advantage for $\log \lambda$ for the non-reversible Markov chains.

References

[1] M. Jerrum, "Mathematical foundations of the Markov chain Monte Carlo method," in *Probabilistic methods for algorithmic discrete mathematics*, ser. Algorithms and Combinatorics. Springer, 1998, vol. 16, pp. 116–165.

[2] P. Diaconis, "The Markov chain Monte Carlo revolution," *Bulletin of the American Mathematical Society*, vol. 46, no. 2, pp. 179–205, 2009.

[3] A. Olshevsky and J. N. Tsitsiklis, "Convergence speed in distributed consensus and averaging," *SIAM J. Control Optim.*, vol. 48, no. 1, pp. 33–55, 2009. [Online]. Available: http://dx.doi.org/10.1137/060678324

[4] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007.

[5] P. Csermely, A. London, L.-Y. Wu, and B. Uzzi, "Structure and dynamics of core/periphery networks," *Journal of Complex Networks*, vol. 1, no. 2, pp. 93–123, 2013.

[6] B. Gerencsér and J. Hendrickx, "Mixing rate on the cycle with some added connections," *Journal of Theoretical Probability*, 2018, accepted for publication.

[7] D. Levin, Y. Peres, and E. Wilmer, *Markov chains and mixing times*. American Mathematical Society, 2009.

[8] B. Gerencsér, J. Hendrickx, and P. van Dooren, "Spectral gap of Markov chains on a cycle," in *14th annual European Control Conference*, 2015.

[9] B. Gerencsér, "Markov chain mixing time on cycles," *Stoch. Proc. Appl.*, vol. 121, no. 11, pp. 2553–2570, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0304414911001785